From DEPARTMENT OF MEDICINE - SOLNA
Karolinska Institutet, Solna, Sweden

LUNG CANCER IN SWEDEN

– INCIDENCE, DIAGNOSIS AND SURVIVAL

Lukas Löfling

Karolinska Institutet

Solna 2020

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, Stockholm, Sweden

ISBN 978-91-7831-955-8

Lung cancer in Sweden - Incidence, diagnosis and survival

THESIS FOR DOCTORAL DEGREE (PhD)

The thesis will be defended at Rolf Luft Auditorium (L1), Karolinska University Hospital, Solna 4th December 2020, 9:00 am

By

Lukas Löfling

Principal Supervisor:
Shahram Bahmanyar, Associate Professor
Karolinska Institutet, Sweden
Department of Medicine - Solna
Unit of Clinical Epidemiology
Centre for Pharmacoepidemiology

Co-supervisors:
Gunnar Wagenius, Associate Professor
Karolinska University Hospital, Sweden
Cancer Theme

Helle Kieler, Professor
Karolinska Institutet, Sweden
Department of Medicine - Solna
Unit of Clinical Epidemiology
Centre for Pharmacoepidemiology

Opponent:
Odd Terje Brustugun, Associate Professor
Drammen Hospital, Norway
Department of Oncology

Examination Board:
Gisela Helenius, Associate Professor
Örebro University, Sweden
Department of Medical Sciences

Therese Andersson, PhD
Karolinska Institutet, Sweden
Department of Medical Epidemiology and Biostatistics

Fredrik Enlund, Associate Professor
Regional Hospital of Kalmar County, Sweden
Diagnostic Centre

ABSTRACT

Globally, lung cancer is the most commonly diagnosed form of cancer, as well as the number one cancer-related cause of death.

This thesis includes four population-based studies that considered different aspects of lung cancer, covering incidence, diagnosis and survival.

Study I considered the incidence of lung cancer in relation to use of antimuscarinic medications to treat overactive bladder. We identified first-time users of antimuscarinic medications and matched them with individuals not exposed to antimuscarinic medications. Exposed individuals had lower incidence of lung cancer than those unexposed. The inverse association became more pronounced over time from the start of treatment and with the amount of medication from filled prescriptions. Our finding of an inverse association generates hypotheses regarding the prevention of cancer and new treatment strategies for patients with cancer.

In **study II**, we identified individuals with incident non-small cell lung cancer and compared characteristics and survival by smoking status at diagnosis. We found that women, adenocarcinoma, and epidermal growth factor receptor mutation were overrepresented among never-smokers. Furthermore, compared to current smokers, survival was longer for never-smokers. Our findings emphasise the need for an improved understanding of lung cancer among never-smokers that may help to prevent lung cancer and improve survival.

Patients with incident lung cancer were identified in **study III** and matched with individuals free of lung cancer. We investigated patterns of recent use of antibiotics as an indicator of early symptoms of lung cancer. We found that a diagnosis of lung cancer was associated with increased likelihood of recent pre-diagnostic use of antibiotics. The likelihood became more pronounced with the number of filled prescriptions and with proximity to the diagnosis. Our findings further emphasise the importance of ruling out lung cancer following pneumonia treatment.

In **study IV**, we identified patients with incident adenocarcinoma or squamous cell carcinoma of the lung and investigated temporal trends in relative survival. We found that relative survival increased between 1995 and 2016. The increase was most pronounced among women, patients with stage III cancer, patients with adenocarcinoma, and never-smokers. These findings corroborate results from other countries. The increase in relative survival for patients with lung cancer in recent decades can probably be attributed to the improvements in diagnostic procedures and new treatments.

LIST OF SCIENTIFIC PAPERS

I. **Löfling L**, Sundström A, Kieler H, Bahmanyar S, Linder M. Exposure to antimuscarinic medications for treatment of overactive bladder and risk of lung cancer and colon cancer. *Clin Epidemiol* 2019;11:133–143.

II. **Löfling L**, Karimi A, Sandin F, Bahmanyar S, Kieler H, Lambe M, Lamberg K, Wagenius G. Clinical characteristics and survival in non-small cell lung cancer patients by smoking history: A population-based cohort study. *Acta Oncol* 2019;58(11):1618–1627.

III. **Löfling L**, Bahmanyar S, Kieler H, Lambe M, Wagenius G. Antibiotic use prior to a lung cancer diagnosis: A population-based study. (*Manuscript*).

IV. **Löfling L**, Bahmanyar S, Kieler H, Lambe M, Wagenius G. Temporal trends in lung cancer survival: A population-based study. (*Manuscript*).

CONTENTS

LIST OF ABBREVIATIONS

ACE	Angiotensin-converting enzyme
ALK	Anaplastic lymphoma kinase
ARB	Angiotensin receptor blocker
ATC	Anatomical therapeutic chemical
BMI	Body mass index
CDR	Cause of Death Register
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
CT	Computed tomography
DCO	Death certificate only
DDD	Defined daily dose
EBUS-TBNA	Endobronchial ultrasound-guided transbronchial needle aspiration
EGFR	Epidermal growth factor receptor
HR	Hazard ratio
IARC	International Agency for Research on Cancer
ICD	International Classification of Diseases
ICD-O	International Classification of Diseases for Oncology
KRAS	Kirsten rat sarcoma viral oncogene homologue
LCBaSe	Lung Cancer DataBase Sweden
LISA	Longitudinal Integration Database for Health Insurance and Labour Market Studies
MGR	Multi-generation Register
NBHW	National Board of Health and Welfare
NLCR	Swedish National Lung Cancer Register
NPR	National Patient Register
NSCLC	Non-small cell lung cancer
OAB	Overactive bladder
OR	Odds ratio
PD-1	Programmed cell death protein 1
PD-L1	Programmed death ligand 1
PDR	Prescribed Drug Register

PET	Positron emission tomography
PM	Particulate matter
PS	Performance status
RCC	Regional Cancer Centre
RCT	Randomised controlled trial
SBRT	Stereotactic body radiotherapy
SCLC	Small cell lung cancer
SCR	Swedish Cancer Register
TB	Tuberculosis
TKI	Tyrosine kinase inhibitor
TNM	Tumour-node-metastasis
USA	United States of America
VATS	Video-assisted thoracic surgery
VEGF	Vascular endothelial growth factor
WHO	World Health Organization

1 INTRODUCTION

1.1 CANCER

Cancer is a broad group of more than 200 related diseases characterised by uncontrolled cell division (1, 2). Normal cells grow and divide to form new cells as they are needed by the body. When normal cells grow old or become damaged, they die by entering programmed cell death (*apoptosis*) and can be replaced by new cells. Cells may become damaged and mutated, potentially failing to respond to many of the signals that control cell growth and apoptosis. Compared to normal cells, the mutated cells can divide more rapidly and evade apoptosis. With the cell growth advantages of the mutated cells, they can make more replications of themselves than a normal cell can, and their offspring may outperform their non-mutated counterparts in the competition for resources. As more and more cells are created, they form a lump of cells, a tumour. The tumour continues to grow and after a while, it will be of a size that makes it possible to detect it. A hallmark of cancer (malignant tumours) is the capacity to invade surrounding tissue or spread to other organs (metastasis). Benign tumours are non-cancerous tumours that cannot invade surrounding tissue or metastasise. In 2018, more than 60,000 individuals were diagnosed with cancer in Sweden (3). Worldwide, the corresponding number was 18 million (4).

1.2 LUNG CANCER

Globally, lung cancer is the most commonly diagnosed cancer with approximately 2.1 million new cases each year (12% of all newly diagnosed cancers), as well as the number one cancer-related cause of death with approximately 1.8 million deaths each year (18% of cancer-related deaths) (4). In Sweden, approximately 4,000 new lung cancer cases are diagnosed each year with approximately the same number of yearly deaths, making it the fourth most common cancer in women and the seventh most common in men and the number one cancer-related cause of death in Sweden (3). The lung cancer incidence for Swedish men peaked in the 1980s (Figure 1) (5), while the incidence for women in Sweden started to level off in 2010 (Figure 1). Today, the age-standardised lung cancer incidence in Sweden is below 20 per 100,000 person-years for both men and women, making Sweden one of the Western countries with the lowest incidence, especially among men (5, 6).

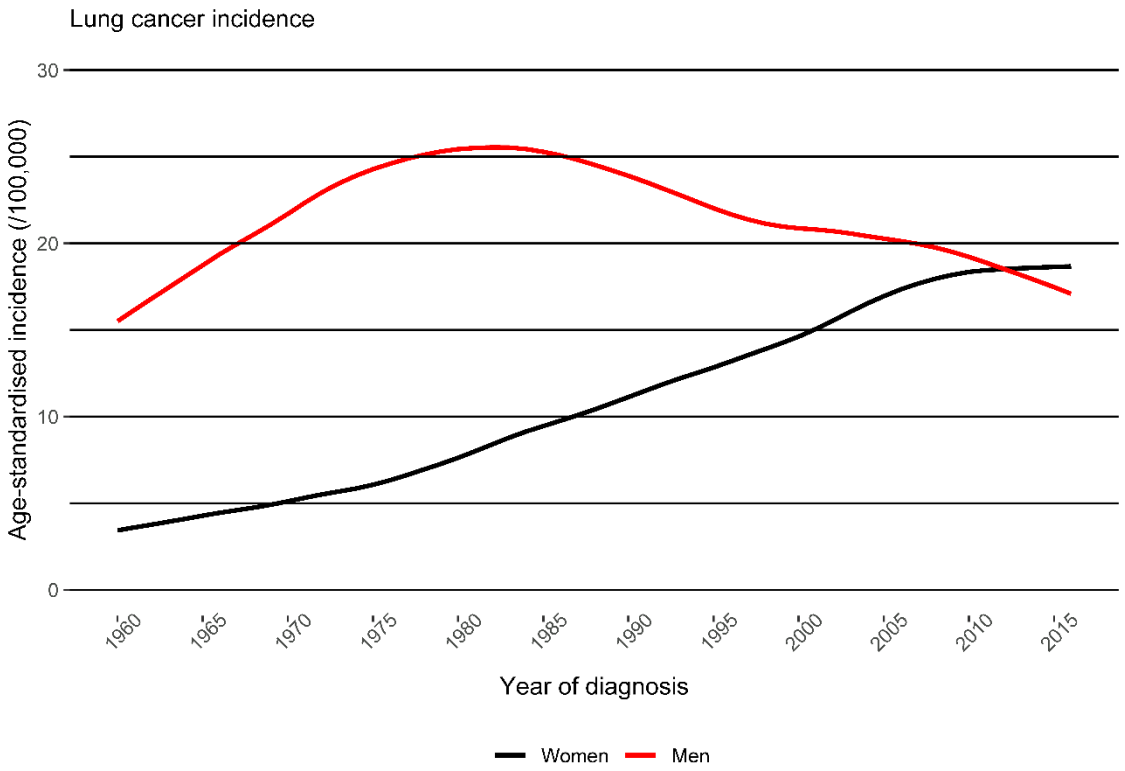

Figure 1. Age-standardised (world standard population) lung cancer incidence in Sweden between 1960 and 2016

1.2.1 Histopathology

Based on histopathology, lung cancer is divided into two main groups: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (7, 8). NSCLC accounts for approximately 80% of all lung cancer cases and will be the main focus of this thesis. Adenocarcinoma and squamous cell carcinoma are the two main forms of NSCLC. The different types of lung cancer arise from different cell types in the epithelium of the lung.

Adenocarcinoma arises from the epithelium in the early stage of the gland (*adenoma*) cells (7-12). Adenocarcinoma of the lungs tends to start in the more peripheral part of the lung (*alveoli*) and is generally characterised by slower growth compared to other types of lung cancer. Adenocarcinoma is overrepresented in never-smokers, women and younger patients. Driver gene alterations, for example, epidermal growth factor receptor (*EGFR*) mutation, Kirsten rat sarcoma viral oncogene homologue (*KRAS*) mutation, and gene fusions of anaplastic lymphoma kinase (*ALK*) or *ROS1*, occur more often in patients with adenocarcinoma compared to other histopathological subtypes of lung cancer. The proportion of patients with adenocarcinoma has increased over time, and today, more than 50% of the patients with lung cancer in Sweden have adenocarcinoma, making it the most common subtype of lung cancer (13).

Squamous cell carcinoma is cancer that begins in squamous cells, often in a main or lobar bronchus of the lung; it tends to be locally aggressive and may invade local structures (7-12). Compared to adenocarcinoma, oncogene driver alterations are rare in squamous cell carcinoma. However, a mutation in the *TP53* gene is more common in squamous cell carcinoma. Squamous cell carcinoma is overrepresented among ever-smokers, men and older patients. Currently, squamous cell carcinoma is the second most common subtype of lung cancer; approximately 20% of patients with lung cancer in Sweden have squamous cell carcinoma, and the proportion has been fairly stable in recent years (13).

SCLC is a centrally located tumour type that starts in neuroendocrine-cell precursors of the lung (7, 8, 10). SCLC is the subtype of lung cancer with the most pronounced association with smoking, almost 100% of patients with SCLC are current or former smokers. Oncogene driver alterations are rare in SCLC. Approximately 15% of the patients with lung cancer in Sweden have SCLC, with an indication of a decreasing trend in the last few years (13).

1.2.2 Stage

Cancer stage at diagnosis is a well-established indicator of disease severity and prognosis (14-17). In this thesis, stage at diagnosis is based on the tumour-node-metastasis (TNM) classification system by the American Joint Committee on Cancer, using the fourth edition from 1992, the fifth edition from 1997, the sixth edition from 2002, and the seventh edition from 2010 onwards. The T-descriptor refers to the size and localisation of the tumour, the N-descriptor describes the involvement of the lymph nodes, and the M-descriptor refers to the presence of metastasis. The specific characteristics of the descriptors within the seventh edition are presented in Table 1. The changes between editions tend to mainly involve refined divisions of the T-descriptor. One substantial change from the sixth to the seventh edition was that pleural effusion was reclassified from T4 to M1. The T-, N- and M-descriptors are grouped in different combinations to ascertain the cancer stage, ranging from I to IV. The different combinations for the seventh edition are presented in Table 2.

Table 1. Description of the tumour-node-metastasis (TNM) descriptors for lung cancer in the seventh edition of the TNM classification system by the American Joint Committee on Cancer

Descriptor	Description
T: Primary tumour size and location	
TX	Primary tumour cannot be assessed or tumour proven by the presence of malignant cells in sputum or bronchial washings but not visualised by imaging or bronchoscopy
T0	No evidence of primary tumour
Tis	Carcinoma *in situ*
T1	Tumour <3 cm in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopy evidence of invasion more proximal than the lobar bronchus (i.e. not in the main bronchus)
T1a	Tumour <2 cm in greatest dimension
T1b	Tumour >2 cm but <3 cm in greatest dimension
T2	Tumour >3 cm but <7 cm or tumour with any of the following features (T2 tumours with these features are classified T2a if <5 cm): • Involves main bronchus, >2 cm distal to the carina • Invades visceral pleura • Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumour >3 cm but <5 cm in greatest dimension
T2b	Tumour >5 cm but <7 cm in greatest dimension
T3	Tumour >7 cm or one that directly invades any of the following: • Chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium • Tumour in the main bronchus <2 cm distal to the carina but without involvement of the carina • Associated atelectasis or obstructive pneumonitis of the entire lung • Separate tumour nodule(s) in the same lobe

Table 1. Description of the tumour-node-metastasis (TNM) descriptors for lung cancer in the seventh edition of the TNM classification system by the American Joint Committee on Cancer

Descriptor	Description
T4	Tumour of any size that invades any of the following: • Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, oesophagus, vertebral body, carina • Separate tumour nodule(s) in a different ipsilateral lobe to that of the primary tumour
N: Regional lymph nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
M: Distant metastases	
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumour nodule(s) in a contralateral lobe, tumour with pleural nodules or malignant pleural (or pericardial) effusion
M1b	Distant metastasis in extrathoracic organ(s)

Table 2. Stage according to the seventh edition of the tumour-node-metastasis (TNM) classification system for lung cancer by the American Joint Committee on Cancer

T/M	N0	N1	N2	N3
T1a	IA	IIA	IIIA	IIIB
T1b	IA	IIA	IIIA	IIIB
T2a	IB	IIA	IIIA	IIIB
T2b	IIA	IIB	IIIA	IIIB
T3	IIB	IIIA	IIIA	IIIB
T4	IIIA	IIIA	IIIB	IIIB
M (any T)	IV	IV	IV	IV

1.2.3 Performance status (PS)

PS is a standard measure for the general health status of patients with cancer and of how the disease impacts the daily life of the patient (18). According to the World Health Organization (WHO), PS is graded from 0 to 5 as described in Table 3.

Table 3. World Health Organization performance status

Grade	Description
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g. light housework, office work
2	Ambulatory and capable of all self-care but unable to carry out any work activities; up and about more than 50% of waking hours
3	Capable of only limited self-care; confined to bed or chair more than 50% of waking hours
4	Completely disabled; cannot carry out any self-care; totally confined to bed or chair
5	Dead

1.2.4 Diagnosis

In addition to concluding that it is lung cancer, the diagnostic procedure aims to ascertain the histopathology, genetic alterations, cancer stage and PS (19). All these components of the diagnosis are of great importance when deciding on treatment strategy.

Diagnostic procedures

Since 2016, there has been a standardised course of care for lung cancer in Sweden that aims to standardise the course of care throughout the whole of Sweden (20). It includes a set of recommended investigative activities and recommended maximum lead times from a well-founded suspicion of lung cancer to start of treatment. However, there is still a long way to go until the recommended lead times are reached for >80% of the patients (Table 4) (21).

Table 4. Lead time from a well-founded suspicion of lung cancer to start of treatment

	Maximum[1]	Median (percentage within maximum)[2]		
Treatment		**2016**	**2017**	**2018**
Surgery	44	73 (20)	72 (13)	71 (10)
Radiotherapy	44	48 (46)	69 (27)	70 (27)
Chemotherapy/medical treatment	40	40 (51)	43 (44)	49 (37)
Palliative symptom relief care	30	26 (61)	28 (53)	29 (52)

Data presented: median lead time in days (percentage within the maximum lead time as recommended in the standardised course of care)

[1] Maximum lead time in days from a well-founded suspicion of lung cancer to start of treatment as recommended in the standardised course of care.

[2] Reported lead time in days from a well-founded suspicion of lung cancer to start of treatment.

Newly presented respiratory symptoms that are persistent for longer than six weeks among smokers or former smokers over the age of 40, unexplainable chest or shoulder pain or haemoptysis should cause suspicion of lung cancer (19, 22). For never-smokers, these symptoms should cause suspicion of lung cancer in the absence of other explanations. Individuals with suspected lung cancer should be referred to undergo a lung X-ray or

computed tomography (CT) of the lungs. Pathological findings from an initial lung X-ray or CT, recurrent haemoptysis, metastatic findings that give rise to suspicion of lung cancer, superior vena cava obstruction or vocal cord paralysis should give rise to a well-founded suspicion of lung cancer, and these patients should be referred to a specialist clinic for further examination. However, when symptoms are starting to present, the lung cancer has usually metastasised, which is the case for the majority of newly diagnosed cases (13). Typically, early detection is often an accidental finding after radiology was indicated for other reasons.

All patients should undergo a clinical examination including auscultation and percussion of the lungs, auscultation of the heart, measurement of the blood pressure, palpation of the supraclavicular lymph nodes and palpation of the abdomen (19). A correct clinical examination is an important first step when assessing the cancer stage.

Examination with CT in combination with positron emission tomography (PET), required for a more precise diagnosis concerning metastases and tumour size, is recommended for patients considered for curative treatment (13, 19). The use of PET-CT started in 2007 and has increased over time; in 2016, almost 100% of the patients considered for curative treatment had undergone an examination with PET-CT as a part of their diagnostic procedure. Due to the low risk of metastases, patients with stage IA are not included in the recommendations regarding PET-CT.

Bronchoscopy is a part of the routine diagnostic procedure and aims to assess the endobronchial expansion and to verify the diagnosis by taking a biopsy (pathological diagnosis) and/or sampling cells by scraping or brushing (cytological diagnosis) (19). Conventional bronchoscopy works best for central tumours, while if the tumour is peripheral, one can perform CT- or ultrasound-guided transthoracic fine needle aspiration or core biopsy. The collected tissue can also be used for molecular profiling of the tumour. For patients with squamous cell carcinoma, testing is recommended for at least the programmed cell death protein ligand 1 (*PD-L1*), while for patients with other forms of NSCLC, in addition to testing for *PD-LI*, testing is recommended for different driver gene alterations, for example, *EGFR* mutation, *KRAS* mutation, and rearrangement of *ALK* and *ROS1*.

In patients considered for curative treatment, bronchoscopy can be combined with endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) (19). EBUS-TBNA is the recommended bronchoscopy method for sampling from mediastinal lymph nodes and should be performed after a PET-CT has been performed. Mediastinal sampling is recommended for these patients in the case of abnormal lymph nodes detected on a PET-CT or radiologically enlarged lymph nodes (even if PET-negative).

1.2.5 Aetiology and risk factors

Tobacco use

Tobacco smoking is the main risk factor for lung cancer and has been estimated to cause >80% of all lung cancer cases in high-income countries, with a lower corresponding proportion in many low- and middle-income countries (8, 23, 24). The prevalence of smoking peaked earlier among men and in high-income countries compared to women and people in low- and middle-income countries. Because of the long latency period, incidence trends of lung cancer reflect smoking prevalence 20–30 years earlier (25, 26). The International Agency for Research on Cancer (IARC) has identified more than 50 different carcinogenic substances present in tobacco smoke (27, 28). The mechanism of these substances involves, for example, creation of free radicals that cause DNA damage in lung cells, subsequently resulting in cancer development.

The association between tobacco smoking and lung cancer has been reported in several epidemiological studies since the early 1950s (29, 30). The excess risk of lung cancer comparing regular smokers to never-smokers is approximately tenfold, with a more pronounced magnitude with increasing daily cigarette consumption and an indication of a more pronounced magnitude for women compared to men (31).

Although the evidence is strongest for first-hand smoking, there is also an observed association between second-hand smoking and lung cancer (32). It was estimated in a meta-analysis that compared to not living with a smoker, the excess risk of lung cancer in never-smokers who lived with a smoker was 24%.

The use of electronic cigarettes as a tobacco-free alternative to regular cigarettes is increasing (33, 34). Compared to regular cigarettes, electronic cigarettes produce fewer toxic and carcinogenic substances. However, the high temperature in the vapour still generates many toxic substances that are classified as carcinogenic to humans. Today, the long-term carcinogenic effect of electronic cigarette use remains unknown.

Studies that investigated the association between use of smokeless tobacco and lung cancer have reported inconclusive results, with an indication of a positive association reported in studies from the United States of America (USA) and indications of an inverse association in studies from Norway and Sweden (35). These differences may be attributed to differences in the composition and consumption habits of the smokeless tobacco products.

Radon

Radon, a naturally occurring radioactive noble gas resulting from the decay of uranium-238, was established as the first environmental lung carcinogen (36, 37). Radon decays in the lungs and emits alpha radiation, resulting in DNA damage and subsequent tumour development. It is estimated that indoor radon exposure contributes to approximately 10% of new cases of lung cancer (19). However, Swedish case-control studies have estimated an increased risk among smokers only, suggesting a synergistic effect between radiation and tobacco smoking (38, 39).

Asbestos

Asbestos, which consists of silicate mineral fibres, was first used commercially in the late 19th century, with a peak between the 1950s and 1970s in high-income countries (40-43). Asbestos is the most common occupational cause of lung cancer, WHO has estimated that approximately 125 million people worldwide are exposed to asbestos in an occupational setting. In the 1930s, the link between asbestos and lung cancer was hypothesised. However, it was not until the 1950s that asbestos became recognised as a lung carcinogen after observational studies on industrial workers in the United Kingdom.

Air pollution

The combined evidence suggests a positive association between outdoor air pollution and lung cancer, and the IARC has classified outdoor air pollution as a lung carcinogen in humans (44). A meta-analysis including 17 studies from nine European countries found an association between exposure to particulate matter (PM) with a diameter smaller than 10 μm (PM_{10}) and lung cancer, with an increased risk of approximately 20% per every 10 $\mu g/m^3$ increase in PM_{10} concentration (45).

Infections

Lung cancer has mainly been associated with infections of the respiratory system. In a meta-analysis including 41 studies, tuberculosis (TB) infection was associated with increased incidence of lung cancer (46). The association was most pronounced between one and five years after the TB infection and declined over time. However, a positive association was still observed 20 years after the TB infection. Moreover, *Chlamydia pneumoniae* has been associated with an increased risk of a lung cancer diagnosis (47). All six studies in a review from 2005 reported a positive association. However, it is important to consider that respiratory infections are the most common differential diagnoses of lung cancer. This can introduce bias leading to a stronger association due to the fact that early symptoms of lung cancer can be misdiagnosed as a respiratory infection.

Three major mechanisms of infectious agent-mediated cancer are often suggested (48). The first is the induction of a chronic inflammatory environment as a consequence of a continuing immune response to a persistent infection. The second mechanism is virus-induced transformation, which causes a change in growth, phenotype or replication of cells. The third mechanism is the chronic suppression of the immune system by the infectious agent.

Pharmacological treatments

Different pharmacological treatments (e.g. angiotensin-converting enzyme [ACE] inhibitors, antibiotics and aspirin) have been assessed for association with lung cancer incidence in observational studies.

ACE inhibitors, which are medications used in the treatment of hypertension, have been associated with an increased risk of different cancers, including lung cancer. A large cohort study in 2018 with data on almost one million individuals from the United Kingdom treated for hypertension found that compared to angiotensin receptor blocker (ARB) use, use of ACE inhibitors was associated with a 14% increased risk of lung cancer (49).

Two studies using data from the United Kingdom estimated increased odds (approximately 30%) of lung cancer associated with antibiotic use (50, 51). It is hypothesised that the association between antibiotic use and an increased risk of lung cancer is mediated by suppression of the normal anti-tumour response in the cells or a change in the composition of the human microbiota (52).

A meta-analysis of 15 studies estimated an inverse association between use of aspirin and lung cancer incidence (approximately 15% lower risk) (53). The magnitude of the inverse association was more pronounced for case-control studies than for cohort studies, suggesting a possible role of recall bias.

Diet

The topic of diet and lung cancer risk is controversial with conflicting results (54). Antioxidants are found in many fruits and vegetables and act as electron donors which neutralise free radicals that can cause DNA damage and promote tumour development. There is some evidence from epidemiological studies that a diet rich in fruits and vegetables is associated with lower incidence of lung cancer, while it has been suggested that a high intake of red or processed meat can increase the risk of lung cancer. If the association is real, it may be explained by the formation of nitrosamines when preparing the meat at high temperatures, a substance suspected to be carcinogenic in humans.

Family history

Several studies have assessed the role of family history in lung cancer. A case-control study with pooled data on 24,380 patients with lung cancer and 23,399 individuals free of lung cancer from The International Lung Cancer Consortium estimated that having a first degree relative with lung cancer was associated with a 1.5-fold increase in the odds of lung cancer (55). The magnitude of the effect size for the association was more pronounced in the Asian population, for patients diagnosed with lung cancer before the age of 50, and for siblings compared to parents. The stronger association for siblings is probably attributed to shared environmental childhood exposure or an indication that recessive genes are involved.

Genetic factors

Mutation in the *TP53* gene is one of the most frequent mutations in patients with lung cancer, more than 50% of whom have a mutation in the *TP53* gene. Mutation frequency increases with the duration of smoking and is more common in squamous cell carcinoma than in adenocarcinoma (8). Other relatively common alterations of genes involved in lung cancer are mutations in *EGFR* and *KRAS*, and rearrangements in *ALK* and *ROS1*. The prevalence of these alterations is less common among patients with non-adenocarcinoma. Among patients with adenocarcinoma, 10–15% have an *EGFR* mutation, 25–40% have a mutation in *KRAS*, 2–10% have a rearrangement of the *ALK* gene, and 1% have a rearrangement of the *ROS1* gene.

Sex

The increased risk of cancer in men has been observed for many cancers, including lung cancer, with decreasing difference over time (56-58). The observed discrepancy in incidence between men and women is often explained by differences in environmental exposures, for example, tobacco smoking, radon, asbestos, diet, and air pollution. Furthermore, innate risk factors have been proposed as potential reasons for the sex-based discrepancy, for example, the total number of cells, sex hormones, and immunological factors (11, 58, 59). Today, lung cancer is more common among women than men in Sweden, most probably due to changes in smoking habits in recent decades (13).

1.2.6 Treatments

The treatment strategy for lung cancer depends mainly on four factors: histopathology, genetic alterations, cancer stage, and PS (19, 60).

Surgery

The first successful restriction of a lung tumour by removing an entire lung (pulmectomy) was performed in 1933 by Graham and Singer (61). Today, a surgical restriction is the preferred curative treatment for patients with NSCLC with stages I and II but can also be considered for some patients with stage IIIA (19, 60). In patients with stage IIIB, surgery is mostly considered not feasible because of an invasion of the mediastinal structure of the vertebrae (T4) or substantial involvement of lymph nodes (N3). In general, there is no place for surgery in patients with stage IV cancer. However, it can be considered for selected cases, for example, in the case of a single metastasis in the brain or the adrenal glands. The type of surgical restriction is dependent on the location, size and spread of the tumour and the PS of the patient. Currently, about 80% of all surgical restrictions of patients with lung cancer are lobectomies or bilobectomies, where one or two lung lobes are removed respectively. Surgical restriction of lung tumours is performed either with open thoracotomy or by video-assisted thoracic surgery (VATS). Proper staging, to exclude mediastinal and extrathoracic metastases, and assessment of the patients' PS, lung function and comorbidities, should be done before a decision on surgery can be made.

Platinum-based chemotherapy can be given to patients with NSCLC who undergo a complete tumour restriction, either as neoadjuvant chemotherapy (i.e. chemotherapy is given prior to the surgery) or as adjuvant chemotherapy (i.e. chemotherapy is given after surgery) (19). The reason for giving neoadjuvant chemotherapy is mainly to reduce the tumour size prior to surgery, while the rationale for giving adjuvant chemotherapy is to eliminate micrometastases to reduce the risk of relapse. The benefit of adjuvant chemotherapy after surgery is well-established for patients with stages IB–IIIA (60, 62, 63). A pooled analysis published in 2008 by Pignong *et al* with 4,584 patients from five randomised controlled trials (RCTs) estimated longer overall survival associated with adjuvant chemotherapy after surgery compared to surgery alone (hazard ratio [HR] = 0.89, 95% confidence interval [CI]: 0.82–0.96) (62). For the patients with stage IA, an indication of shorter survival was estimated for the patients receiving adjuvant chemotherapy compared to surgery alone (HR = 1.40, 95% CI: 0.95–2.06). Another meta-analysis published in 2010 with 34 RCTs that assessed the same question estimated longer overall five-year survival associated with adjuvant treatment (64% and 60% respectively; HR = 0.86, 95% CI: 0.81–0.92), also indicating shorter survival for patients with stage IA (HR = 1.19, 95% CI: 0.84–1.68) (63). Neoadjuvant treatment has also been assessed, and a meta-analysis with data from ten RCTs including 2,188 patients with NSCLC in stages IB–IIIA reported prolonged survival associated with receiving neoadjuvant chemotherapy compared to surgery alone (HR = 0.89, 95% CI: 0.81–0.98) (64). A study from the USA that compared survival between neoadjuvant and adjuvant chemotherapy in 35,134 patients with NSCLC in stages II–III estimated lower hazard of death associated with adjuvant therapy (stage II – median overall survival: 81 and 67 months respectively; HR =

0.75, 95% CI: 0.65–0.88, and stage III – median overall survival: 49 and 42 months respectively; HR = 0.80, 95% CI: 0.70–0.91) (65).

Based on current knowledge, patients with NSCLC, except stage IA, who undergo a complete tumour restriction should be offered platinum-based chemotherapy in addition to surgery (19). The standard treatment approach in Sweden is a platinum-based combination given as adjuvant chemotherapy in four cycles that should be initiated within eight weeks after surgery.

Radiotherapy

Patients with NSCLC in stages I and II who are inoperable because of, for example, comorbidity, poor PS or an unfavourable location of the tumour, should be treated with stereotactic body radiotherapy (SBRT) (19, 60). Compared to conventional radiotherapy, SBRT uses high doses of radiation delivered to a very precise area. The first RCT comparing SBRT and conventional radiotherapy among inoperable patients with stage I NSCLC, published in 2016, reported that there was no estimated difference in either progression-free survival (HR = 0.85, 95% CI: 0.52–1.36) or overall survival (HR = 0.75, 95% CI: 0.43–1.30) (66). However, SBRT was associated with a better quality of life and less toxicity. An observational study from 2018 that included more than 20,000 patients with NSCLC in stage I from the USA reported longer overall survival associated with SBRT compared to conventional radiotherapy (median overall survival: 39 and 28 months respectively; HR = 0.73, 95% CI: 0.69–0.77) (67).

Combined radio-chemotherapy is the standard treatment for curative intent in patients with inoperable NSCLC in stage III (19, 68). The combination can be given sequentially (starting with chemotherapy and subsequent radiotherapy) or concomitantly (chemotherapy and radiotherapy are given at the same time). The aim of sequential therapy is mainly to reduce the frequency of micrometastases and to reduce the tumour size prior to radiotherapy, while the aim of concomitant therapy is mainly to achieve a high locoregional control. In terms of survival, concomitant treatment is considered superior to sequential treatment (68-70). In Sweden, the current recommendation for combined radio-chemotherapy is to give concomitant radio-chemotherapy with three cycles of a platinum-based regime, and radiotherapy is added from the second cycle onwards (19). However, concomitant therapy is also associated with a higher occurrence of severe oesophageal toxicity. Therefore, patients with poor PS or comorbidity should receive sequential therapy.

Chemotherapy and immunotherapy

Chemotherapy can be given with either curative or palliative intent. When given with curative intent, it is given in combination with surgery or radiotherapy, as described above. This section will focus on chemotherapy given with palliative intent.

Chemotherapy and immunotherapy are presented together in this section as their indications highly overlap. In general, they are indicated for patients with late-stage (IIIB–IV) NSCLC

without targetable alterations, such as *EGFR* mutation or *ALK* or *ROS1* rearrangements (19, 71).

In recent decades, platinum-based chemotherapy has been the cornerstone in the palliative treatment of patients with late-stage NSCLC (19, 72, 73). However, because of the limited efficacy and significant toxicity in the early days of chemotherapy, the use of chemotherapy was a controversial issue until the 1990s. A meta-analysis from 1995 with eight RCTs including a total of 778 patients with late-stage NSCLC estimated that treatment with platinum-based chemotherapy was associated a 27% lower hazard of death compared to treatment with best supportive care alone, equivalent to an increase in one-year survival from 5% to 15% (73). These results were later confirmed by an updated meta-analysis published in 2008, which included 16 RCTs and a total of 2,714 patients (HR = 0.77, 95% CI: 0.71–0.83) (72). In terms of survival, there is an indication that carboplatin-based regimes are associated with a slightly lower survival compared to cisplatin (HR = 1.07, 95% CI: 0.99–1.15) (74). Furthermore, a higher response rate was found among patients treated with cisplatin compared to carboplatin. However, thanks to their more favourable toxicological profile, carboplatin-based regimes are becoming more common in palliative care (19, 74). The optimal number of treatment cycles with platinum-based chemotherapy in first-line treatment for patients with late-stage NSCLC has been investigated in different meta-analyses. There is no strong evidence for prolonged overall survival associated with more than four cycles of chemotherapy (75-77). However, there is evidence for prolonged progression-free survival, as well as an indication of more adverse events and more impaired quality of life in patients treated with more than four cycles.

The involvement of the immune system in cancer development has been known for a long time. The programmed cell death protein 1 (*PD-1*) is a receptor expressed on the surface of activated T-cells and its ligand, *PD-L1*, is expressed on the surface of different immune cells (78). To avoid chronic autoimmune response, the interaction between *PD-1* and *PD-L1* inhibits the T-cell response and ensures that the immune response mediated by T-cells is activated only when appropriate. *PD-L1* is often overexpressed on tumour cells, resulting in pronounced inhibition of the immune system by the tumour. Immunotherapy is the newest choice in the treatment of patients with lung cancer and has recently led to changes in recommendations of treatment for patients with late-stage NSCLC. Immunotherapy acts by binding to the *PD-1* receptor, thereby preventing the interaction between *PD-1* and *PD-L1*; this will result in activation of the immune system that can identify and eliminate the tumour.

Immunotherapy as monotherapy in treatment-naïve patients with late-stage NSCLC and *PD-L1* expression ≥50% has been reported to be associated with longer survival when compared to the patients treated with platinum-based chemotherapy (79). Furthermore, combination therapy of platinum-based chemotherapy and immunotherapy in treatment-naïve patients with late-stage NSCLC has been associated with longer survival when compared to platinum-based chemotherapy alone (80, 81). The prolonged survival associated with the combination therapy was independent of *PD-L1* expression degree.

Based on current knowledge, combination therapy of platinum-based chemotherapy and immunotherapy is recommended for patients with late-stage NSCLC without targetable alterations such as *EGFR* mutation or rearrangements of the ALK or *ROS1* genes and with a WHO PS <2 (19). It is recommended for patients in this group with expression of *PD-L1* ≥50% that immunotherapy is given as monotherapy, while patients with a WHO PS of 2 or patients with increased risk of complications from immunotherapy should be given platinum-based chemotherapy as monotherapy. Upon progression for patients undergoing platinum-based chemotherapy as monotherapy, one should consider changing to treatment with immunotherapy as monotherapy. If there is progression for patients on immunotherapy given as monotherapy, switching to platinum-based chemotherapy is an option, while for patients who progress after combination therapy, switching to docetaxel or pemetrexed can be an alternative.

Tyrosine kinase inhibitors (TKIs)

EGFR is a transmembrane growth factor receptor that serves as a mediator of extracellular growth factors (82). The receptor is involved in cell growth, proliferation and apoptosis. A mutation in the tyrosine kinase domain of the receptor can result in increased downstream activity of pro-survival functions of the cell, such as increased cell proliferation and inhibition of apoptosis. The *EGFR*-TKIs bind to the mutated tyrosine kinase domain and prevent the downstream signalling, resulting in inhibition of the cell proliferation and induction of cell death by the normal apoptotic pathway. Several RCTs have reported prolonged progression-free survival, a higher response rate and a more distinct improvement in the quality of life associated with *EGFR*-TKI compared to platinum-based chemotherapy in treatment-naïve patients with late-stage NSCLC with *EGFR* mutation (83-87). However, clear prolongation in overall survival has been more difficult to find. The combination of *EGFR*-TKI and platinum-based chemotherapy compared to *EGFR*-TKI alone in this patient group was associated with prolonged overall survival and higher occurrence of adverse events (88).

A meta-analysis with five RCTs that included a total of 660 patients with *EGFR*-mutated NSCLC in stages I–IIIA estimated that treatment with adjuvant *EGFR*-TKI after surgery was associated with longer disease-free survival (i.e. time from randomisation to recurrence of tumour or death) compared to surgery alone (HR = 0.48, 95% CI: 0.36–0.65) (89). An indication of longer overall survival associated with adjuvant *EGFR*-TKI treatment was also found (HR = 0.72, 95% CI: 0.49–1.06).

ALK is a transmembrane tyrosine kinase receptor (90). Most alterations of *ALK* genes result in fusion genes, for example, with *EML4*, resulting in the *EML4-ALK* fusion gene where the extracellular and intramembranous parts of the *ALK* are replaced with *EML4*. *ALK* fusions can lead to alteration of the kinase activity, and subsequently an increase in cell proliferation and inhibition of apoptosis. *ALK*-TKIs act by blocking the tyrosine kinase activity, resulting in a reduction of the cell proliferation and induction of the intrinsic apoptosis. It has been reported that compared to platinum-based chemotherapy, *ALK*-TKIs are associated with prolonged progression-free survival, a higher objective response rate and a greater improvement in the quality of life in treatment-naïve patients with *ALK*-positive late-stage

NSCLC (91, 92). Similar to *EGFR*-TKIs, a prolongation in overall survival has been difficult to find, probably due to the high number of patients switching treatment upon disease progression.

Based on current knowledge, patients with late-stage NSCLC should undergo treatment with an *EGFR*-TKI if they have an *EGFR* mutation and with an *ALK*-TKI if they have a rearrangement in the *ALK* gene (19). Furthermore, it is recommended that patients with the rare *ROS1* rearrangement should undergo treatment with an *ALK*-TKI. Upon disease progression on TKIs, one may first consider other TKIs for the same alteration. However, platinum-based chemotherapy can be an option and should in those situations be given as described above.

Anti-angiogenesis

Angiogenesis refers to the formation of new blood vessels that are responsible for the supply of oxygen and nutrition to the tumour (2). Solid tumours cannot grow larger than 2–3 mm or metastasise without forming their own blood vessels, making tumour angiogenesis a hallmark of cancer. Vascular endothelial growth factor (*VEGF*) is the key mediator of angiogenesis, as it binds to the *VEGF* receptors. Upregulation of *VEGF* is a response to hypoxia, which is common in cancer, resulting in angiogenesis.

Bevacizumab, a monoclonal antibody, binds to *VEGF* and thereby prevents it from binding to the receptor, resulting in inhibition of angiogenesis (93). In patients with late-stage adenocarcinoma, not eligible for a TKI therapy, treatment with bevacizumab in combination with platinum-based chemotherapy has been associated with prolonged progression-free survival and an increased response rate compared to platinum-based chemotherapy alone (94-97).

1.2.7 Survival

Survival for patients diagnosed with lung cancer is generally low (98). In most countries, between 2010 and 2014, the overall five-year relative survival was estimated at between 10% and 20%. In Sweden, it was estimated at 20%, and the highest estimated overall five-year relative survival was in Japan at 33%.

Prognostic factors

A prognostic factor is a factor that is associated with the natural history of the disease, namely progression-free survival or overall survival independent of the treatment given (99, 100). Prognostic factors in lung cancer encompass both tumour-related factors, for example, cancer stage at diagnosis (101), histopathology (102) and genetic alterations (103-105), and patient-related factors, for example, PS (102, 106), sex (102, 107, 108), age (102), socioeconomic status (109) and smoking status (102, 110-112).

Cancer stage

Cancer stage at diagnosis is probably the most important prognostic factor in patients with lung cancer. For the eighth edition of the TNM classification, data on more than 90,000 patients diagnosed with lung cancer between 1999 and 2010 from 16 countries was analysed (101). The reported five-year survival varied between 92% for patients in the earliest stage and 0% for patients with multiple distant metastases.

Histopathology

The prognostic value of histopathology is unclear due to conflicting results. Five of 31 studies in a comprehensive review that assessed histopathology as a prognostic factor in patients with NSCLC reported a difference in prognosis by histopathology (102). Of these five studies, longer survival was associated with adenocarcinoma in four and with squamous cell carcinoma in one.

Genetic alterations

The prognostic value of several genetic alterations has been studied, for example, it has been suggested that the wild-type of the *TP53*, *EGFR* and *KRAS* genes is associated with a more favourable prognosis (103-105). Although a large number of studies have assessed the prognostic impact of *TP53* mutation, the results are inconsistent, mainly because of relatively small study populations. A systematic review and meta-analysis published in 2001 by Steels *et al* which included 67 studies assessing survival by *TP53* mutation status in patients with NSCLC reported that 30 studies identified *TP53* mutation as a negative prognostic factor (103). It was estimated that having a *TP53* mutation was associated with shorter survival of patients compared to those with a wild-type of the *TP53* gene (HR = 1.40, 95% CI: 1.20–1.72). Meert *et al* conducted a meta-analysis of 11 studies assessing the prognostic impact of *EGFR* mutation and reported an HR for death of 1.14 (95% CI: 0.94–1.39) by comparing patients with a mutation to patients with a wild-type (104). Meng *et al* included 41 studies in a meta-analysis assessing the prognostic impact of mutation of the *KRAS* gene and reported shorter survival associated with a mutated *KRAS* gene (HR = 1.45, 95% CI: 1.29–1.62) (105).

PS

Despite the subjective aspects of PS, it is considered one of the strongest prognostic factors in lung cancer (102, 106). The association between PS and survival in patients with lung cancer has been consistently found in studies from different settings and periods. A systematic review with studies published between 2000 and 2010 identified 49 studies assessing the prognostic value of PS, of which 36 found that good PS was associated with longer survival (102). Kawaguchi *et al* analysed data from 26,957 individuals diagnosed with lung cancer in Japan and found a trend of increased hazard of death with increasing WHO PS (i.e. with poorer PS) (106). A WHO PS of 1 was associated with approximately 30% higher hazard of death compared to a WHO PS of 0, while the corresponding estimate for a WHO PS of 4 was 300%.

Sex

There is a large body of evidence supporting longer survival in women compared to men after a diagnosis of lung cancer (102, 107, 108). A systematic review with 45 studies which assessed sex as a prognostic factor reported that 17 of the studies estimated that being a woman was associated with longer survival, while the remaining 28 had not been able to observe a sex-based difference in survival (102). A meta-analysis with 86,800 patients with lung cancer found that being a woman was associated with approximately 20% lower hazard of death compared to men (HR = 0.78, 95% CI: 0.75–0.81) (107). A study from the USA reported longer survival associated with being a woman among untreated patients (HR = 0.85, 95% CI: 0.77–0.93), suggesting that lung cancer has a different natural history in women compared to men (108). The reason for the observed sex-based difference in survival is not fully understood, and different explanations are often proposed and have been discussed in the literature, for example, smoking history, the influence of sex hormones, and differences in tumour biology (11, 108).

Age

The reported results on age as a prognostic factor for lung cancer survival are conflicting. A review published in 2014 which included 39 studies that assessed the prognostic impact of age reported that only four of the studies had estimated an association with age (102). Three of these reported longer survival for older patients, while one reported longer survival for younger patients.

Socioeconomic status

The positive association between socioeconomic status and survival in patients with lung cancer (i.e. higher survival for those with higher socioeconomic status) is observed in different settings, including countries with universal healthcare systems (109). A Swedish study found that a high educational level was associated with longer survival in patients with NSCLC in stages I–II compared to those with a low educational level (HR = 0.86, 95% CI: 0.77–0.92) (113). An association was not found among patients with more advanced stages (stage IIIA: HR = 1.12, 95% CI: 0.99–1.26; stages IIIB–IV: HR = 0.99, 95% CI: 0.95–1.03)

(113). Similar results of an association between socioeconomic group and survival for patients with early-stage lung cancer have been reported from England (114).

Smoking

The prognostic impact of smoking is well-established, and being an ever-smoker has consistently been associated with shorter survival (102, 110-112). A review with six studies which assessed the difference in survival by smoking status reported that all six studies estimated longer survival associated with a shorter or no history of smoking (102). It is estimated that approximately 50% of patients who smoked at diagnosis continued to smoke after diagnosis (115). One hypothesised reason for the survival difference is that tobacco smoke alters the metabolic rate of several chemotherapies by inducing enzymes in the cytochrome P450 system, resulting in lower plasma levels of chemotherapies (116, 117). Other potential explanations discussed in the literature are the impaired wound healing caused by smoking, smoking-associated oxidative stress, immune suppression resulting in cancer progression, nicotine-induced angiogenesis that stimulates and facilitate tumour growth, and post-diagnosis development of smoking-related comorbidity increasing the risk of death (115, 118, 119).

2 OBJECTIVES

The overall objectives of the thesis are to investigate possible factors associated with a subsequent diagnosis of lung cancer and survival after a diagnosis of lung cancer using routinely collected healthcare data from a universal healthcare setting.

The specific objectives of each of the four included studies are:

I. To estimate the association between filling a prescription of antimuscarinic medications used to treat overactive bladder (OAB) and a subsequent diagnosis of lung cancer.

II. To compare clinical characteristics, demographics and survival in patients with NSCLC by smoking status at diagnosis.

III. To examine patterns of recent fillings of prescriptions of antibiotics recommended for the treatment of pneumonia as an indicator of early symptoms of lung cancer.

IV. To delineate temporal trends in lung cancer survival, overall and by important prognostic factors.

3 MATERIALS AND METHODS

3.1 DATA SOURCES

For study I, we acquired data from the Swedish Cancer Register (SCR), the National Patient Register (NPR), the Prescribed Drug Register (PDR), the Cause of Death Register (CDR), and the Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA). In studies II–IV, the data was acquired from the Lung Cancer DataBase Sweden (LCBaSe). Record linkage between the registers was made feasible using the personal identity number, a unique personal identifier given to all permanent residents of Sweden, which is included in all of the registers (120). The registers and databases used as data sources in this thesis are described in detail below.

LCBaSe

The LCBaSe is a research database generated by record linkage between the Swedish National Lung Cancer Register (NLCR) and several other population-based registers in Sweden, namely the SCR, the CDR, the NPR, the PDR, the Multi-generation Register (MGR), the LISA database, and the Swedish Population Register (Figure 2). Each patient with lung cancer in the LCBaSe has been individually matched with up to five individuals free of lung cancer from the general Swedish population. The matching was based on year of birth, sex, and place of residence at the time of the diagnosis of lung cancer. In this thesis, we used data in the LCBaSe from the NLCR, the SCR, the CDR, the NPR, the PDR, the LISA database, and the Swedish Population Register; these registers are described in detail below (121-127).

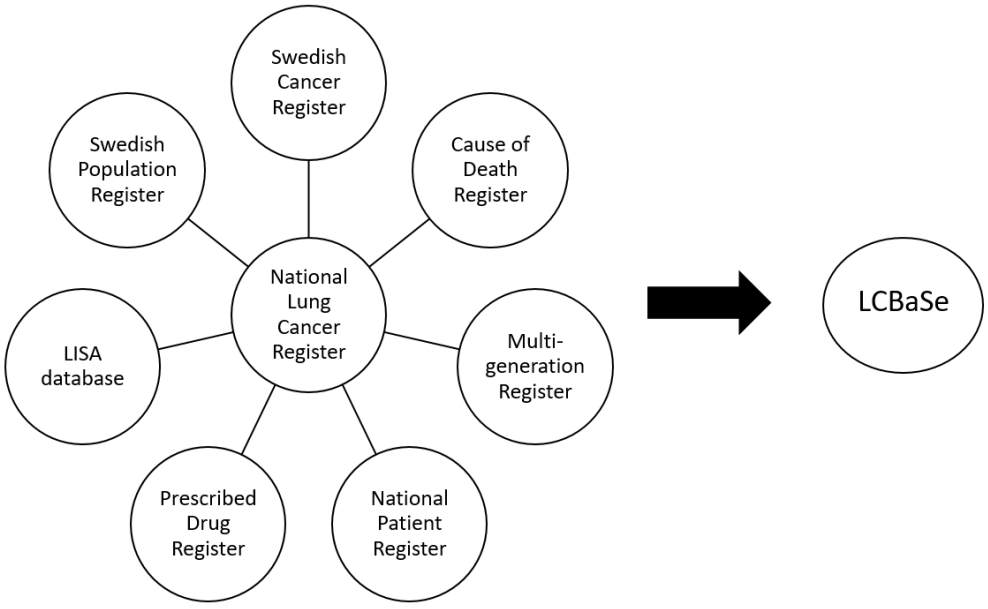

Figure 2. Linkage of registers to generate the Lung Cancer DataBase Sweden (LCBaSe)

NLCR

The NLCR, a nationwide population-based register established in 2002, aims to include all newly diagnosed patients with lung cancer (International Classification of Diseases [ICD] seventh edition [ICD-7] code: 1621 or ICD for oncology third edition [ICD-O-3] code: C34) in Sweden (121, 128, 129). Before 2002, the register was a regional population-based register in the healthcare region of Uppsala-Örebro (approximately two million inhabitants, 20% of the entire Swedish population). Compared to the SCR, where reporting is mandated by law, the NLCR covers more than 95% of newly diagnosed patients with lung cancer in Sweden. The NLCR records individual-level information on, for example, sex, age, smoking status, histopathology, cancer stage, PS, *EGFR* mutation status (2010 onwards), diagnostic procedures, basis of diagnosis, and planned primary treatment. The NLCR does not include tumours discovered at autopsy, that is, there is no inclusion of death certificate only (DCO) cases. The NLCR is held by the Regional Cancer Centre (RCC) in the healthcare region of Uppsala-Örebro.

SCR

The SCR was established in 1958 and records individual-level data (e.g. site of cancer, histopathology, date of diagnosis and basis of diagnosis) on all newly diagnosed malignant tumours in Sweden (122). The cancers are recorded using the latest versions of the ICD codes (ICD-7 [1952–1968], ICD-8 [1969–1986], ICD–9 [1987–1996] and ICD-10 [1997 onwards]) or ICD-O codes (ICD-O-2 [1993–2004] and ICD-O-3 [2005 onwards]), and the historical seventh edition (ICD-7) is always included (130). Reporting to the register is mandated by law. In 1998, it was estimated that the underreporting to the register was approximately 4%, with higher underreporting in the elderly, which is similar to other population-based cancer registers in Europe (131-134). The six RCCs in Sweden perform the registration of incident cancers and annually send their data to the National Board of Health and Welfare (NBHW), where the SCR is held. In this thesis, we did not include data on DCO cases from the SCR.

CDR

The CDR is held at the NBHW and was established in 1961 (123, 135). The register includes individual-level information, from death certificates, on the underlying and contributing causes of death (using the latest version of ICD codes) and the date of death for all deaths occurring among Swedish residents each year. Stillbirths are not included in the CDR. Less than 1% of recorded deaths in the CDR are missing an underlying cause of death (coded as death certificate not received). The overall agreement between the cause of death recorded in the register and relevant information from medical records has been estimated at approximately 80%, and 90% for malignancies, with higher agreement in younger individuals (136). The underlying cause of death is defined in ICD-10 as ‘(a) the disease or injury which initiated the train of morbid events leading directly to death, or (b) the circumstances of the accident or violence which produced the fatal injury’ (137). In the presence of more than one cause of death recorded by the physician on the death certificate, the process to assign an underlying cause is complex. In brief, the physician has to specify the conditions that played

a role in the death and then separate the conditions that directly led to the death from other significant conditions that contributed to the death.

NPR

The NPR is held by the NBHW and was created in 1964 as the Hospital Discharge Register that collected individual-level data on somatic inpatient care, and since 1973, psychiatric inpatient care has been included (124). Due to regional discrepancies in recording practices, national coverage was not achieved until 1987. Since 2001, the register has included information on outpatient visits, with increasing coverage during the first few years. Main and secondary diagnoses are recorded using the latest version of ICD codes. Information on diagnoses given in primary care is not included in the NPR. It is estimated that the national completeness for inpatient care is approximately 99% for somatic and psychiatric discharges, and the validity of the recorded information is approximately 90% (138). However, the completeness is lower (80%) for outpatient care.

PDR

Since July 2005, the PDR has recorded individual-level information on all filled prescriptions from community pharmacies, for example, the active substance (coded according to the anatomical therapeutic chemical [ATC] codes), amount (e.g. the number of tablets or defined daily dose [DDD]), date of prescription, and date when the prescription was filled (125, 139). Initially, information on filled prescriptions is reported by the community pharmacies to the Swedish eHealth Agency and is subsequently transferred to the NBHW, where the PDR is created and held. The completeness and validity of the data are virtually 100%, mainly due to the fact that the registration is almost exclusively automatic. The PDR does not include data on over-the-counter medications or medications used in hospitals, nursing homes or prisons.

LISA database

The sociodemographic information in this thesis was retrieved from the LISA database (126). The database integrates existing data from registers on, for example, education and income. Currently, the LISA database holds information on all permanent residents in Sweden aged 16 years and over for every year since 1990.

Swedish Population Register

The Swedish Population Register covers all permanent residents in Sweden and is held by the Swedish Tax Agency (127). The register contains information on, for example, date of birth, death and migration.

3.2 MAIN STATISTICAL METHODS

3.2.1 Survival analysis

Survival analysis refers to the area of analysis of time-to-event data (140). Subjects are followed over time from a pre-specified time of origin to the occurrence of an event of interest or censoring. Censoring occurs when the event status of a subject is unknown. The studies in this thesis only consider right censoring. Right censoring is that up until a specific time point no event has been registered due to, for example, migration or administrative censoring (i.e. no more data is available). After that time point, although we know that the event could occur, we are not able to determine if or when. Non-informative censoring is assumed in all studies included in this thesis. Non-informative censoring means that the time of censoring is independent of the event time, namely the distribution of censoring times provides no information on the distribution of event times. This thesis considers two types of events for time-to-event data: cancer as an outcome (i.e. cancer incidence) and death (i.e. mortality).

The survival function gives the probability of having a survival time, T, after time t:

$$S\,(t) = P(T > t), 0 < t < \infty$$

The survival function can only take a value between 0 and 1 (0% and 100%) and cannot increase by time (i.e. non-increasing by nature). The hazard function describes the instantaneous rate at time t of experiencing the event of interest, conditional on having survived until time t:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

The main methods in the area of survival analysis applied in this thesis are described in detail below.

Kaplan–Meier method

The Kaplan–Meier method is a non-parametric method used to estimate survival curves, namely one is not assuming anything regarding the shape of the underlying hazard function (140, 141). The estimator is obtained by:

$$S(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j}\right)$$

where n_j is the number of individuals at risk at time t_j and d_j is the number of failures or events (e.g. diagnoses of cancer or deaths) at time t_j. The survival probability is estimated at every time an event occurs and is calculated as the number of individuals surviving that time interval divided by the number of individuals at risk at the start of that time interval. Individuals who have been censored (e.g. emigrated) before that time interval are not considered to be at risk. The total probability of surviving until a specific time is calculated by multiplying all the probabilities of survival preceding that specific time interval.

In RCTs, the Kaplan–Meier method is the main method used to analyse and present time-to-event data since the randomisation makes it reasonable to assume that the populations are the same apart from the exposure (142). However, in observational studies, to control for potential confounders, it is often necessary to adjust the analyses.

Application in study II

The Kaplan–Meier method was used in study II to estimate unadjusted survival curves among patients diagnosed with lung cancer. This was carried out by smoking status overall and in subgroups of cancer stage at diagnosis.

Cox proportional hazard model

Currently, the Cox proportional hazard model is the most common method used to analyse time-to-event data in observational studies (140, 143). The Cox model is formulated as:

$$\lambda(t) = \lambda_0(t) * \exp(\beta_1 x_1 + \cdots + \beta_n x_n)$$

where $\lambda(t)$ is the hazard function; $\lambda_0(t)$ is the baseline hazard function, namely when all covariates are at their pre-specified baseline levels; x_n are the covariates; and β_n are their estimated coefficients. In the model, only the βs are estimated, while the $\lambda_0(t)$ is not estimated; therefore, the Cox model is a semi-parametric model, namely one can estimate the relative effect (i.e. HR) but not the absolute effect. The basic assumptions of the Cox model are that the hazards for the compared populations are proportional over time and that the censoring is non-informative (i.e. censoring times are unrelated to the event times). In this thesis, the proportional hazards assumption was tested by the introduction of an interaction term between the exposure and the follow-up time (underlying time scale) or with scaled Schoenfeld residuals.

Application in studies I and II

In study I, a Cox proportional hazard model was used to study the incidence of lung cancer by exposure to antimuscarinic medications. In this case, the hazard represents the incidence rate and the HR is the incidence rate ratio between exposed and unexposed individuals.

In study II, a Cox proportional hazard model was used to study mortality among patients with lung cancer in relation to their smoking status at diagnosis. When studying mortality, the hazard represents the mortality rate and the HR is to be interpreted as the mortality rate ratio between the groups.

Relative survival

Relative survival is an estimation of the net survival (i.e. probability of survival in the absence of other causes of death) and is defined as the ratio of the all-cause survival of the patients to the expected survival (144-147):

$$R(t) = \frac{S(t)}{S^*(t)}$$

where $S(t)$ is the overall survival in the patient cohort and $S^*(t)$ is the expected survival at time t after the diagnosis. The expected survival is calculated for a whole population, derived from life tables, which are often used as a proxy for the survival that the diseased would have experienced had they not been diseased. In this thesis, this is the survival for patients with lung cancer had they been free of lung cancer. The main advantage of using relative survival instead of cause-specific survival is that the validity of the estimates does not rely on the validity of the reported cause of death.

Application in study IV

In study IV, relative survival and temporal trends in relative survival for patients with lung cancer were estimated. This was carried out overall and in different subgroups defined by relevant prognostic factors (i.e. sex, histopathology, cancer stage, and smoking status).

3.2.2 Logistic regression

Logistic regression models are common and useful in epidemiological research when the outcome variable is binary (e.g. 0/1, no/yes, no diagnosis/diagnosis). Logistic regression is used to model the probability of a certain event given a set of covariates. For a binary outcome variable, this is formulated as:

$$Odds = \frac{p}{1-p} = exp\,(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)$$

where p is the probability of the outcome, x_n are the covariates, and β_n are their estimated coefficients.

Application in study III

In study III, a logistic regression model was used to estimate the association between a diagnosis of lung cancer and a recent history of filled prescriptions of antibiotics.

4 SUMMARY OF STUDIES

4.1 STUDY I

Background: Preclinical studies have found that antimuscarinic substances inhibit cell proliferation in lung cancer.

Study population: Individuals with a first-time filled prescription of an antimuscarinic medication used to treat OAB between 2006 and 2012 (index date) as recorded in the PDR (exposed). Each exposed individual was matched (based on year of birth, sex, and place of residence at the index date) with unexposed individuals (i.e. not exposed before the index date) from the general population.

Study design: Population-based cohort study.

Setting: Sweden.

Exposure: Antimuscarinic medications indicated for the treatment of OAB.

Main outcome: A diagnosis of lung cancer (ICD-7:1621, 163; ICD-O-3: C34, C39). Colon cancer was also included in the study but will not be discussed in this thesis.

Statistical analyses: A Cox proportional hazard model was used to estimate HR for the association between exposure and a diagnosis of lung cancer. The unexposed individuals were the reference group. Follow-up started on the index date and ended on the date of one of the following events: a diagnosis of lung cancer (i.e. event of interest), death, emigration or administrative censoring (31 December 2013), whichever occurred first. The proportional hazard assumption was tested by introducing an interaction term between exposure and follow-up time. If the proportional hazard assumption was not met, HRs were estimated separately for different intervals of the follow-up time. The data was analysed using an intention-to-treat approach, namely no consideration was taken of change in exposure status after the index date.

Results: The study included 164,000 individuals classified as exposed to antimuscarinic medications and 1,446,472 unexposed individuals. The estimated incidence rate difference was −44.9 (95% CI: −53.3, −36.5) per 100,000 person-years. The negative incidence rate difference indicates that there is a lower incidence rate for the exposed individuals compared to the unexposed individuals. Also, the HR below one from the Cox model indicates that being exposed was associated with lower incidence of lung cancer (Table 5). The magnitude of the HR for the inverse association became more pronounced with longer time since the index date and an indication of a pronounced association for the highest group of cumulative DDD (≥365 DDDs). However, we did not find a difference in the effect size by sex. The point estimate for the HR became slightly more pronounced when applying a lag time (i.e. no outcomes considered within a specific period) (Table 6).

Table 5. HR and 95% CI for the association between exposure to antimuscarinic medications and lung cancer, Swedish Cancer Register and the Prescribed Drug Register, 2006–2012. Time since index date (i.e. date of incidence prescription of a study medication for the exposed individuals and the corresponding date for the matched unexposed individuals) was used as the underlying time scale

Group		n[1]	HR (95% CI)[2]	
Overall				
Unexposed		8,394	1	(ref)
Exposed		659		
	<1 year	234	0.86	(0.75−0.98)
	1–4 years	347	0.63	(0.56–0.70)
	≥4 years	78	0.43	(0.34–0.55)
Women				
Unexposed		3,823	1	(ref)
Exposed		323		
	<1 year	106	0.81	(0.66–0.99)
	1–4 years	179	0.64	(0.54–0.74)
	≥4 years	38	0.37	(0.26–0.51)
Men				
Unexposed		4,571	1	(ref)
Exposed		336		
	<1 year	128	0.90	(0.75–1.08)
	1–4 years	168	0.62	(0.53–0.72)
	≥4 years	40	0.51	(0.37–0.70)
Cumulative DDD[3]				
	≤90	340	1	(ref)
	91–180	103	0.96	(0.76–1.20)
	181–364	83	1.01	(0.79–1.29)
	≥365	133	0.82	(0.67–1.00)

Abbreviations: Hazard ratio (HR), Confidence interval (CI), Defined daily dose (DDD).

[1] Number of individuals with an event of interest, i.e. a diagnosis of lung cancer.

[2] Adjusted for year of birth, sex, place of residence, income, educational level, and a proxy variable for smoking (i.e. a diagnosis of chronic obstructive pulmonary disease or a filled prescription for a smoking cessation medication within five years prior to the index date, i.e. date of incidence prescription of a study medication for the exposed individuals and the corresponding date for the matched unexposed individuals).

[3] Includes exposed individuals only.

Table 6. HR and 95% CI for the association between exposure to antimuscarinic medications and lung cancer, Swedish Cancer Register and the Prescribed Drug Register, 2006–2012, when applying a lag time period directly after index date (i.e. no events of interest are considered within this time period). Time since end of lag time (i.e. six and 12 months respectively after index date, i.e. date of incidence prescription of a study medication for the exposed individuals and the corresponding date for the matched unexposed individuals) was used as the underlying time scale

Group		n[1]	HR (95% CI)[2]	
6 months lag				
Unexposed		7,226	1	(ref)
Exposed		530		
	<1 year	176	0.70	(0.60–0.82)
	1–4 years	311	0.64	(0.57–0.72)
	≥4 years	43	0.33	(0.24–0.45)
12 months lag				
Unexposed		6,107	1	(ref)
Exposed		425		
	<1 year	139	0.62	(0.52–0.74)
	1–4 years	258	0.61	(0.54–0.70)
	≥4 years	28	0.30	(0.21–0.45)

Abbreviations: Hazard ratio (HR), Confidence interval (CI).

[1] Number of individuals with an event of interest, i.e. a diagnosis of lung cancer.

[2] Adjusted for year of birth, sex, place of residence, income, educational level, and a proxy variable for smoking (i.e. a diagnosis of chronic obstructive pulmonary disease or a filled prescription for a smoking cessation medication within five years prior to the index date, i.e. date of incidence prescription of a study medication for the exposed individuals and the corresponding date for the matched unexposed individuals).

Discussion: Our finding of an inverse association between exposure to antimuscarinic medications and incidence of lung cancer was in accordance with results of a study investigating the same association using data from Denmark (148). The more pronounced inverse association with longer follow-up time could indicate that it takes time from the start of the exposure until we can see a biological effect of the exposure on cancer incidence. The indication of a pronounced magnitude of the inverse association for those in the highest group of cumulative DDD (≥365 DDDs) compared to the lower groups is an indication of a dose–response relationship between exposure and outcome. In line with the study from Denmark, we did not find any evidence in support of a sex-based difference in the effect of antimuscarinic medications on lung cancer incidence (148).

Conclusion: We found an inverse association between exposure to antimuscarinic medications, used in the treatment of OAB, and a diagnosis of lung cancer. The magnitude of the effect size became more pronounced with longer time since the start of treatment and with a higher cumulative dose. These findings generate hypotheses regarding the prevention of cancer and potentially new treatment strategies in cancer.

4.2 STUDY II

Background: Previous epidemiological studies on differences in patient characteristics and survival by smoking status in lung cancer have been limited by selective data and small study populations.

Study population: Individuals with a first-time diagnosis of NSCLC between 2002 and 2016 as recorded in the NLCR.

Study design: Population-based cohort study.

Setting: Sweden.

Exposure: The patients were classified based on self-reported smoking status at diagnosis: current smokers (smoked at diagnosis or stopped <1 year before), former smokers (stopped ≥1 year before diagnosis), and never-smokers (never smoked on a regular basis).

Main outcome: Lung cancer-specific mortality (ICD-7: 1621, ICD-10: C34).

Statistical analyses: The Kaplan–Meier method was used to estimate unadjusted lung cancer-specific survival curves. A Cox proportional hazard model was used to estimate HR for lung cancer-specific mortality, with the current smokers as the reference group. Follow-up started on the date of diagnosis and ended on the date of one of the following events: death due to lung cancer (i.e. event of interest), death due to other causes, emigration or administrative censoring (31 December 2016), whichever occurred first. The proportional hazard assumption was tested using the scaled Schoenfeld residuals. If the proportional hazard assumption was not fulfilled, HRs were estimated in separate intervals of the follow-up time by the introduction of an interaction term between the smoking status and the follow-up time.

Results: We included 41,262 patients with NSCLC, of whom 43% were current smokers, 44% former smokers, 11% never-smokers, and data on smoking history was not available for 2%. Compared to current smokers, a higher median age at diagnosis and a larger proportion of patients aged ≤50 years were observed for never-smokers. Moreover, overrepresentation of women, adenocarcinoma, *EGFR* mutation, and stage IV was found among never-smokers. In addition, we found that being a never-smoker was associated with longer survival in the first years after the diagnosis of lung cancer (Figure 3, Table 7). The estimated longer survival associated with being a never-smoker was consistent over cancer stage and sex.

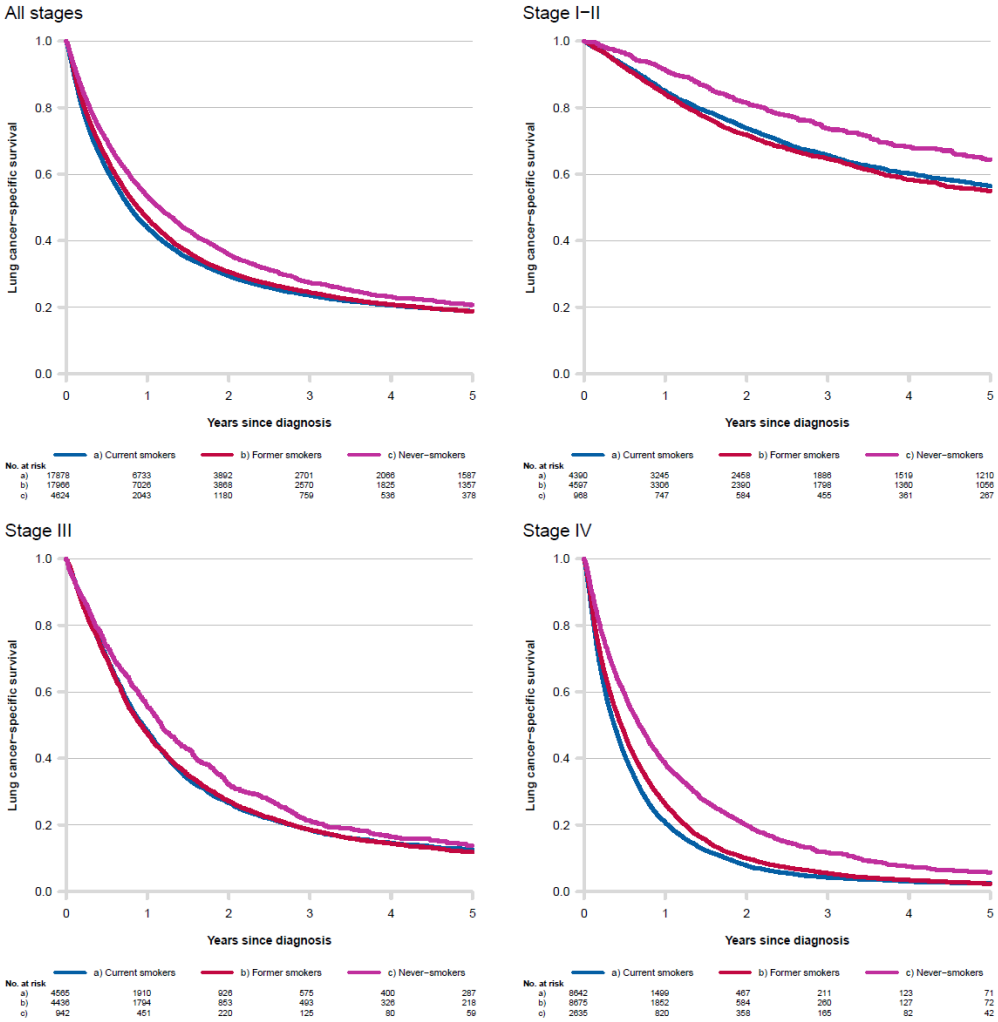

Figure 3. Lung cancer-specific survival by smoking status at diagnosis among patients diagnosed with non-small cell lung cancer, Lung Cancer DataBase Sweden, 2002–2016. Current smokers (blue), former smokers (red) and never-smokers (purple). Smoking status was based on self-reported information at diagnosis: current smokers (smoked at diagnosis or stopped <1 year before), former smokers (stopped ≥1 year before diagnosis), and never-smokers (never smoked on a regular basis)

Table 7. HR and 95% CI for the association between smoking status[1] at diagnosis and lung cancer-specific mortality among patients diagnosed with non-small cell lung cancer, Lung Cancer DataBase Sweden, 2002–2016. Time since diagnosis was used as the underlying time scale.

	Stage at diagnosis					
	I–II		III		IV	
	HR[2] (95% CI)		HR[2] (95% CI)		HR[2] (95% CI)	
Overall						
Current smokers	1	(ref)		(ref)	1	(ref)
Former smokers	0.95	(0.88–1.01)	0.94	(0.89–0.99)	0.90	(0.87–0.93)
Never-smokers	0.81	(0.71–0.93)				
≤0.5 years since diagnosis			0.78	(0.67–0.90)	0.65	(0.61–0.70)
>0.5, ≤1 years since diagnosis			0.70	(0.59–0.83)	0.69	(0.62–0.76)
>1, ≤2 years since diagnosis			0.88	(0.75–1.05)	0.63	(0.56–0.71)
>2 years since diagnosis			1.01	(0.83–1.22)	0.91	(0.77–1.06)
Men						
Current smokers	1	(ref)	1	(ref)	1	(ref)
Former smokers	0.94	(0.86–1.04)	0.97	(0.91–1.04)	0.87	(0.83–0.91)
Never-smokers	0.80	(0.64–1.01)				
≤0.5 years since diagnosis			0.69	(0.53–0.90)	0.65	(0.58–0.72)
>0.5, ≤1 years since diagnosis			0.56	(0.41–0.77)	0.59	(0.50–0.70)
>1, ≤2 years since diagnosis			0.86	(0.65–1.12)	0.55	(0.45–0.67)
>2 years since diagnosis			1.16	(0.85–1.59)	1.02	(0.79–1.32)
Women						
Current smokers	1	(ref)	1	(ref)	1	(ref)
Former smokers	0.94	(0.84–1.05)	0.88	(0.82–0.96)	0.93	(0.88–0.98)
Never-smokers	0.84	(0.71–0.99)				
≤0.5 years since diagnosis			0.83	(0.69–1.01)	0.64	(0.58–0.70)
>0.5, ≤1 years since diagnosis			0.78	(0.63–0.97)	0.79	(0.70–0.90)
>1, ≤2 years since diagnosis			0.91	(0.73–1.12)	0.71	(0.61–0.82)
>2 years since diagnosis			0.92	(0.72–1.18)	0.87	(0.71–1.07)

Abbreviations: Hazard ratio (HR), Confidence interval (CI).

[1] Smoking status was based on self-reported information at diagnosis: current smokers (smoked at diagnosis or stopped <1 year before), former smokers (stopped ≥1 year before diagnosis), and never-smokers (never smoked on a regular basis).

[2] Adjusted for sex (in the overall analysis), age, histopathology, World Health Organization performance status, primary planned treatment, Charlson Comorbidity Index, and educational level.

Discussion: The findings regarding higher median age and overrepresentation of women, adenocarcinoma, and patients with *EGFR* mutation among never-smokers confirm those of previous studies and corroborate the suggestion that lung cancer in never-smokers may be considered a different disease compared to smoking-associated lung cancer (9, 10, 110-112, 149-165). However, the higher proportion of younger patients (≤50 years) among never-smokers compared to smokers has not been described frequently and is an indication of a stronger hereditary effect for lung cancer in never-smokers compared to lung cancer in smokers (55, 166). Moreover, we found that stage IV cancer was more common in never-smokers; however, whether smoking status is associated with cancer stage remains unclear, with conflicting results in the available literature (112, 152, 153, 155, 156, 158-160). A systematic review published in 2019 that investigated barriers of early diagnosis of lung cancer identified low awareness of lung cancer symptoms among both physicians and patients as a barrier (167). It is reasonable to consider the awareness of lung cancer as a

possible reason for presenting symptoms to be lower among never-smokers, and that this may explain the higher proportion of patients with stage IV cancer.

Our finding of longer survival for never-smokers compared to smokers is in accordance with those of other studies (102, 110-112). It has been suggested that tobacco smoke alters the metabolic rate of several chemotherapies by inducing enzymes in the cytochrome P450 system, resulting in lower plasma levels of cytostatic agents and subsequently leading to lower survival for smokers compared to never-smokers (116, 117). Other suggested reasons for the difference in survival are impaired wound healing caused by smoking, smoking-associated oxidative stress, immune suppression resulting in cancer progression, nicotine-induced angiogenesis that stimulates and facilitates tumour growth, and post-diagnosis development of smoking-related comorbidities increasing the risk of death (115, 118, 119).

Conclusion: The observed differences in age and sex distribution, histopathology, *EGFR* mutation, cancer stage, and survival between smokers and never-smokers in this study emphasise the need for an improved understanding of non-tobacco-associated lung cancer that may help to prevent lung cancer and improve survival.

4.3 STUDY III

Background: Little is known about patterns of pre-diagnostic use of antibiotics as an indicator of early symptoms of lung cancer.

Study population: Individuals with a first-time diagnosis of lung cancer between 2009 and 2016 as recorded in the NLCR (cases). Each case was individually matched with individuals free of lung cancer from the general population (controls) using a risk set sampling approach (i.e. concurrent sampling) at the date of the diagnosis for the cases (index date). The matching factors were year of birth, sex, and place of residence at the index date.

Study design: Population-based case-control study.

Setting: Sweden.

Exposure: Recent history (i.e. within three years prior to the index date) of filled prescriptions of antibiotics recommended for the treatment of pneumonia.

Main outcomes: A diagnosis of lung cancer (ICD-7: 1621, ICD-O-3: C34).

Statistical analyses: A logistic regression model was used to estimate the odds ratio (OR) for the association between a diagnosis of lung cancer and a history of recent pre-diagnostic filled prescriptions of antibiotics recommended for the treatment of pneumonia.

Results: A total of 27,017 cases and 129,355 controls were included. Of the cases, 50% had filled at least one prescription, 25% had a history of repeated fillings, and 7% had four or more fillings, with a maximum of 63 treatment cycles. The corresponding proportions for the controls were 33%, 13% and 4% respectively, with a maximum of 78 treatment cycles. Compared to the individuals free of lung cancer, the likelihood of a recent history of at least one filled prescription was higher for the patients with lung cancer, NSCLC overall (OR: 1.83, 95% CI: 1.77–1.88), squamous cell carcinoma (OR: 2.18, 95% CI: 2.05–2.32), adenocarcinoma (OR: 1.70, 95% CI: 1.64–1.77), and SCLC (OR: 1.92, 95% CI: 1.78–2.06). The magnitude of the effect size became more pronounced with an increasing number of filled prescriptions (Figure 4) and with proximity to the diagnosis (Figure 5). However, the magnitude of the effect size did not differ by sex or educational level and became attenuated with increasing age (Table 8).

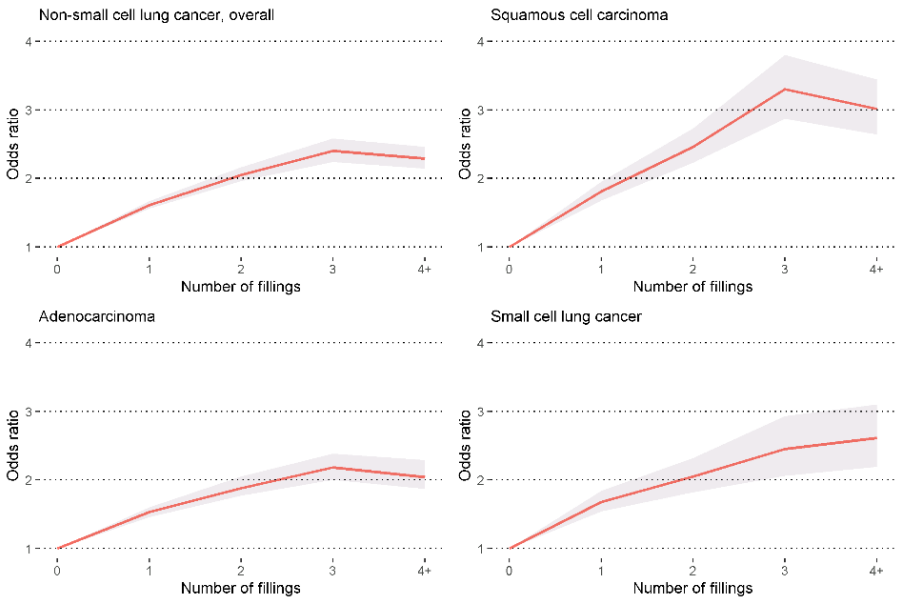

Figure 4. Adjusted odds ratios (solid line) and 95% confidence intervals (shaded area) for the association between a diagnosis of lung cancer and the number of recently filled prescriptions of antibiotics recommended for the treatment of pneumonia, Lung Cancer DataBase Sweden, 2009–2016. The odds ratios were adjusted for year of birth, sex, place of residence, educational level, previous diagnosis of chronic obstructive pulmonary disease, previous use of antibiotics recommended for the treatment of pneumonia, and history of any cancer. The observation period was three years before diagnosis of lung cancer

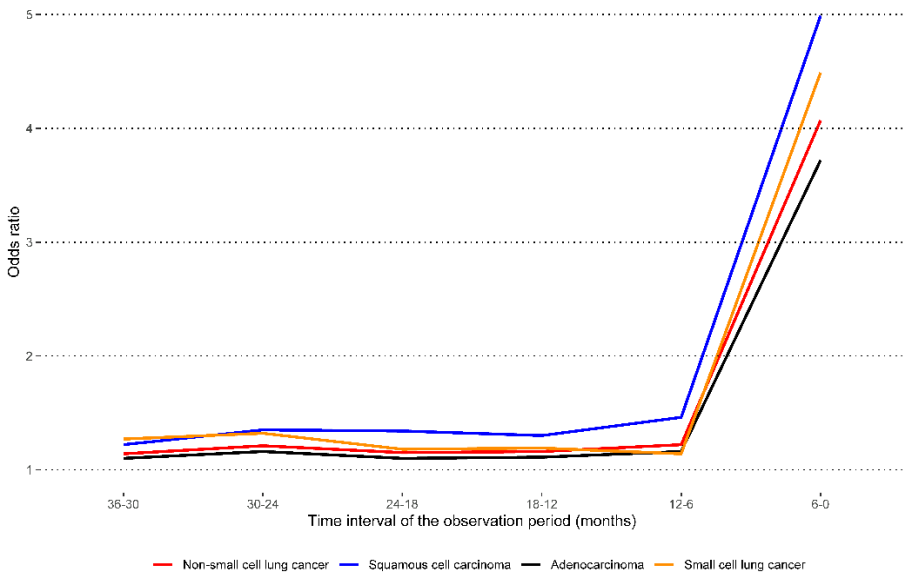

Figure 5. Adjusted odds ratios for the association between a diagnosis of lung cancer and a history of at least one filled prescription of antibiotics recommended for the treatment of pneumonia in different intervals of the period prior to the date of lung cancer diagnosis, Lung Cancer DataBase Sweden, 2009–2016. The odds ratios were adjusted for year of birth, sex, place of residence, educational level, previous diagnosis of chronic obstructive pulmonary disease, previous use of antibiotics recommended for the treatment of pneumonia, and history of any cancer

Table 8. OR and 95% CI for the association between a diagnosis of lung cancer and a recent history of at least one filled prescription of antibiotics recommended for the treatment of pneumonia, Lung Cancer DataBase Sweden, 2009–2016

		Exposed individuals[1]		OR (95% CI)			
		Cases	Controls	Unadjusted		Adjusted[2]	
NSCLC, overall							
Sex							
	Men	5,553	17,382	2.00	(1.92–2.08)	1.80	(1.73–1.88)
	Women	5,949	19,090	2.08	(2.00–2.17)	1.86	(1.78–1.94)
Age at diagnosis (years)							
	<50	257	798	2.26	(1.90–2.71)	2.20	(1.84–2.63)
	50–59	1,136	3,493	2.31	(2.11–2.52)	2.15	(1.97–2.35)
	60–69	4,120	13,131	2.08	(1.99–2.18)	1.91	(1.82–2.00)
	70–79	4,277	13,661	1.96	(1.88–2.05)	1.72	(1.64–1.80)
	≥80	1,712	5,389	1.94	(1.81–2.08)	1.68	(1.56–1.81)
Educational level[3]							
	Low	4,447	11,473	2.06	(1.98–2.16)	1.77	(1.69–1.86)
	Middle	4,941	14,800	2.12	(2.03–2.21)	1.89	(1.81–1.97)
	High	1,970	9,814	1.93	(1.81–2.06)	1.81	(1.69–1.93)
	Missing	144	385	1.90	(1.51–2.39)	1.68	(1.33–2.13)
Squamous cell carcinoma							
Sex							
	Men	1,747	4,977	2.46	(2.31–2.63)	2.13	(2.00–2.28)
	Women	1,219	3,441	2.65	(2.48–2.84)	2.26	(2.10–2.43)
Age at diagnosis (years)							
	<50	26	67	2.97	(2.46–3.59)	2.76	(2.28–3.34)
	50–59	250	599	2.94	(2.65–3.26)	2.63	(2.37–2.93)
	60–69	957	2,696	2.64	(2.46–2.84)	2.33	(2.16–2.51)
	70–79	1,225	3,522	2.45	(2.28–2.62)	2.07	(1.93–2.22)
	≥80	508	1,534	2.40	(2.20–2.61)	2.00	(1.84–2.19)
Educational level[3]							
	Low	1,246	2,906	2.55	(2.38–2.73)	2.11	(1.97–2.27)
	Middle	1,264	3,300	2.64	(2.46–2.82)	2.26	(2.11–2.43)
	High	413	2,118	2.42	(2.22–2.63)	2.17	(1.99–2.37)
	Missing	43	94	2.34	(1.85–2.97)	2.00	(1.57–2.55)
Adenocarcinoma							
Sex							
	Men	3,045	9,980	1.80	(1.72–1.88)	1.65	(1.57–1.73)
	Women	3,980	13,319	1.94	(1.85–2.02)	1.75	(1.67–1.83)
Age at diagnosis (years)							
	<50	198	629	2.16	(1.81–2.59)	2.11	(1.76–2.53)
	50–59	747	2,455	2.14	(1.96–2.34)	2.01	(1.84–2.20)
	60–69	2,667	8,815	1.92	(1.83–2.02)	1.78	(1.69–1.88)
	70–79	2,482	8,342	1.78	(1.69–1.87)	1.58	(1.50–1.67)
	≥80	931	3,058	1.74	(1.62–1.88)	1.53	(1.42–1.65)
Educational level[3]							
	Low	2,582	7,023	1.89	(1.79–1.98)	1.64	(1.56–1.73)
	Middle	3,024	9,546	1.95	(1.86–2.05)	1.76	(1.68–1.85)
	High	1,341	6,495	1.79	(1.67–1.91)	1.69	(1.58–1.81)
	Missing	78	235	1.73	(1.37–2.18)	1.56	(1.23–1.98)
SCLC							
Sex							
	Men	887	2,746	2.11	(1.96–2.28)	1.89	(1.74–2.04)
	Women	1,095	3,350	2.21	(2.05–2.38)	1.95	(1.80–2.10)
Age at diagnosis (years)							

Table 8. OR and 95% CI for the association between a diagnosis of lung cancer and a recent history of at least one filled prescription of antibiotics recommended for the treatment of pneumonia, Lung Cancer DataBase Sweden, 2009–2016

		Exposed individuals[1]		OR (95% CI)			
		Cases	Controls	Unadjusted		Adjusted[2]	
	<50	46	120	2.40	(1.98–2.89)	2.30	(1.90–2.79)
	50–59	223	636	2.43	(2.19–2.72)	2.25	(2.01–2.51)
	60–69	727	2,331	2.20	(2.03–2.39)	1.99	(1.84–2.16)
	70–79	757	2,322	2.07	(1.92–2.24)	1.80	(1.66–1.95)
	≥80	229	687	2.05	(1.86–2.26)	1.76	(1.59–1.94)
Educational level[3]							
	Low	818	1,850	2.20	(2.03–2.38)	1.86	(1.72–2.02)
	Middle	864	2,521	2.25	(2.09–2.44)	1.98	(1.83–2.15)
	High	277	1,680	2.05	(1.87–2.25)	1.89	(1.72–2.08)
	Missing	23	45	2.01	(1.59–2.56)	1.76	(1.38–2.26)

Abbreviations: Odds ratio (OR), Confidence interval (CI), Non-small cell lung cancer (NSCLC), Small cell lung cancer (SCLC).

[1] Individuals with at least one filled prescription of antibiotics recommended for the treatment of pneumonia as recorded in the Prescribed Drug Register within three years before the index date (i.e. the date of lung cancer diagnosis and the corresponding date for the matched individuals free of lung cancer).
[2] Adjusted for year of birth, sex, place of residence, educational level, previous diagnosis of chronic obstructive pulmonary disease, previous use of antibiotics recommended for the treatment of pneumonia, and history of any cancer.
[3] Highest attained educational level the year before the index date categorised by years of formal education: ≤9 (low, mandatory), 10–12 (middle, upper secondary), and ≥13 (high, post-upper secondary).

For the association between a diagnosis of lung cancer and a recent pre-diagnostic history of repeated fillings, we found no trend in the magnitude of the effect size by stage (Table 9). However, the results indicate that the magnitude of the effect size was slightly more pronounced for patients with stage III disease compared to stage IV disease. In separate assessments, there were no trends in the magnitude of the effect size by the T- and N-descriptors.

Table 9. OR and 95% CI for the association between a diagnosis of lung cancer and a recent history of repeated fillings (≥2 fillings) of antibiotics recommended for the treatment of pneumonia, Lung Cancer DataBase Sweden, 2009–2016

		Exposed individuals[1]		OR (95% CI)			
		Cases	Controls	Unadjusted		Adjusted[2]	
NSCLC, overall							
Overall		5,746	14,489	2.57	(2.48–2.67)	2.18	(2.10–2.27)
Stage at diagnosis[3]							
	I–II	1,576	3,660	2.94	(2.74–3.17)	2.28	(2.11–2.47)
	III	1,320	2,996	3.04	(2.81–3.30)	2.54	(2.33–2.77)
	IV	2,796	7,692	2.25	(2.14–2.37)	2.02	(1.91–2.13)
	Missing	54	141	2.45	(1.69–3.57)	2.10	(1.39–3.18)
T-descriptor[3]							
	1	1,311	3,131	2.81	(2.60–3.03)	2.29	(2.12–2.48)
	2	1,594	4,169	2.41	(2.26–2.58)	2.03	(1.89–2.17)
	3	950	2,337	2.63	(2.42–2.87)	2.17	(1.99–2.38)
	4	1,794	4,589	2.55	(2.40–2.71)	2.17	(2.04–2.32)
	Missing	97	263	2.29	(1.79–2.91)	1.87	(1.45–2.40)
N-descriptor[3]							
	0	2,272	5,659	2.53	(2.39–2.68)	2.08	(1.96–2.21)
	1	419	1,156	2.34	(2.06–2.66)	1.93	(1.69–2.20)
	2	1,659	4,190	2.66	(2.50–2.84)	2.25	(2.10–2.40)
	3	1,251	3,080	2.67	(2.48–2.87)	2.29	(2.13–2.47)
	Missing	145	404	2.15	(1.75–2.63)	1.73	(1.40–2.13)
Squamous cell carcinoma							
Overall		15,877	3,291	3.48	(3.23–3.75)	2.75	(2.53–2.99)
Stage at diagnosis[3]							
	I–II	480	962	3.61	(3.14–4.14)	2.53	(2.17–2.95)
	III	514	1,021	3.83	(3.35–4.39)	3.14	(2.71–3.65)
	IV	574	1,264	3.14	(2.78–3.54)	2.62	(2.29–3.00)
	Missing	19	44	3.45	(1.75–6.83)	3.41	(1.58–7.36)
T-descriptor[3]							
	1	256	473	3.96	(3.57–4.40)	3.03	(2.72–3.38)
	2	501	1,012	3.28	(2.99–3.60)	2.59	(2.35–2.85)
	3	290	627	3.52	(3.17–3.91)	2.75	(2.46–3.06)
	4	527	1,151	3.45	(3.16–3.77)	2.77	(2.52–3.04)
	Missing	13	28	3.22	(2.50–4.16)	2.46	(1.89–3.20)
N-descriptor[3]							
	0	610	1,313	3.44	(3.15–3.75)	2.67	(2.44–2.92)
	1	145	296	3.12	(2.71–3.61)	2.44	(2.10–2.82)
	2	490	1,007	3.60	(3.28–3.94)	2.86	(2.60–3.15)
	3	306	585	3.68	(3.33–4.07)	2.97	(2.68–3.30)
	Missing	36	90	2.88	(2.33–3.57)	2.20	(1.77–2.75)
Adenocarcinoma							
Overall		3,419	9,324	2.28	(2.18–2.39)	2.00	(1.90–2.10)
Stage at diagnosis[3]							
	I–II	968	2,391	2.68	(2.44–2.93)	2.17	(1.97–2.40)
	III	598	1,575	2.44	(2.18–2.73)	2.07	(1.83–2.33)
	IV	1,826	5,278	2.08	(1.96–2.2)	1.90	(1.78–2.04)
	Missing	27	80	2.05	(1.24–3.40)	1.80	(1.03–3.13)
T-descriptor[3]							
	1	920	2,341	2.56	(2.37–2.77)	2.14	(1.97–2.32)
	2	914	2,656	2.12	(1.97–2.28)	1.83	(1.70–1.97)
	3	517	1,360	2.28	(2.08–2.50)	1.94	(1.76–2.13)
	4	1,002	2,772	2.23	(2.08–2.39)	1.96	(1.82–2.10)
	Missing	66	195	2.08	(1.63–2.66)	1.73	(1.35–2.24)

Table 9. OR and 95% CI for the association between a diagnosis of lung cancer and a recent history of repeated fillings (≥2 fillings) of antibiotics recommended for the treatment of pneumonia, Lung Cancer DataBase Sweden, 2009–2016

		Exposed individuals[1]		OR (95% CI)			
		Cases	Controls	Unadjusted		Adjusted[2]	
N-descriptor[3]							
	0	1,444	3,766	2.25	(2.11–2.40)	1.89	(1.77–1.98)
	1	216	710	2.04	(1.79–2.33)	1.73	(1.51–1.98)
	2	926	2,562	2.35	(2.19–2.52)	2.03	(1.89–2.19)
	3	749	2,054	2.40	(2.22–2.60)	2.11	(1.95–2.29)
	Missing	84	232	1.88	(1.53–2.31)	1.56	(1.26–1.93)
SCLC							
Overall		1,006	2,446	2.74	(2.51–3.00)	2.25	(2.04–2.49)
Stage at diagnosis[3]							
	I–II	47	116	2.75	(1.83–4.14)	1.99	(1.25–3.18)
	III	312	654	3.52	(2.96–4.18)	2.77	(2.29–3.35)
	IV	629	1,642	2.46	(2.20–2.74)	2.09	(1.85–2.35)
	Missing	18	34	3.79	(1.88–7.69)	3.27	(1.41–7.59)
T-descriptor[3]							
	1	119	264	3.01	(2.67–3.38)	2.41	(2.13–2.72)
	2	170	467	2.59	(2.32–2.89)	2.13	(1.90–2.38)
	3	148	326	2.82	(2.50–3.19)	2.28	(2.02–2.58)
	4	541	1,322	2.74	(2.49–3.01)	2.28	(2.07–2.52)
	Missing	28	67	2.45	(1.90–3.16)	1.96	(1.51–2.55)
N-descriptor[3]							
	0	88	306	2.65	(2.38–2.96)	2.12	(1.89–2.38)
	1	49	111	2.46	(2.10–2.87)	1.97	(1.68–2.31)
	2	399	873	2.79	(2.52–3.09)	2.29	(2.06–2.55)
	3	442	1,080	2.80	(2.53–3.10)	2.34	(2.11–2.60)
	Missing	28	76	2.25	(1.81–2.80)	1.77	(1.41–2.21)

Abbreviations: Odds ratio (OR), Confidence interval (CI), Non-small cell lung cancer (NSCLC), Small cell lung cancer (SCLC).

[1] Individuals with repeated (≥2) fillings of prescriptions of antibiotics recommended for treatment of pneumonia as recorded in the Prescribed Drug Register within three years before the index date (i.e. the date of lung cancer diagnosis and the corresponding date for the matched individuals free of lung cancer).

[2] Adjusted for year of birth, sex, place of residence, educational level, previous diagnosis of chronic obstructive pulmonary disease, previous use of antibiotics recommended for treatment of pneumonia, history of any cancer, and the other TNM descriptors for the descriptor specific estimates (e.g. if estimating the effect of exposure in subgroups of T-descriptor, the estimate was adjusted for N- and M-descriptors).

[3] Based on the tumour-node-metastasis (TNM) classification system by the American Joint Committee on Cancer.

Discussion: Infections are common in the area of tumour growth and can present as an initial symptom of a lung tumour. Moreover, tumours can, initially, be difficult to distinguish from infected loci on a lung X-ray (168-170). Consequently, our finding of a positive association between a diagnosis of lung cancer and a recent history of filled prescriptions was not a complete surprise and does not necessarily reflect inappropriate prescribing of antibiotics. However, with as many as 7% of the patients with lung cancer having filled four or more prescriptions (maximum 63) within three years before diagnosis, contrary to clinical guidelines, our findings indicate the absence of proper clinical reassessment and follow-up in some patients after undergoing treatment for pneumonia (171).

We did not find evidence supporting a trend by stage in the magnitude of the effect size for the association between a diagnosis of lung cancer and a history of repeated fillings. The indication of the slightly attenuated magnitude of the effect size for stage IV cancer compared

to stage III cancer may reflect characteristics in patients with stage III disease associated with increased risk of infections or presence of infection-like symptoms, for example, lymph node involvement, spread to an ipsilateral lobe or invasion of heart or central parts of the lung (17). This does not have to be the case for patients with stage IV disease. However, the descriptor-specific estimates, with no trends in the magnitude of the effect size by the T- or N-descriptors, did not bring clarity to the indicated difference between stage III and stage IV. Taken together, our findings do not provide evidence that a history of repeated treatment cycles of antibiotics is related to diagnostic delays, at least not as reflected in increased likelihood of being diagnosed with a larger tumour or more advanced stage.

Conclusion: We found that a diagnosis of lung cancer was associated with approximately a two-fold increase in the likelihood of a recent history of at least one filled antibiotic prescription. The magnitude of the effect size became more pronounced with an increasing number of filled prescriptions and with proximity to the diagnosis, further corroborating the notion that infection represents an early sign of lung cancer. Therefore, repeated antibiotic use may be an indicator of undiagnosed lung cancer. Our findings do not support the suggestion that repeated treatment cycles is associated with a diagnostic delay, as reflected by a more advanced disease at diagnosis. Our findings further underscore the importance to rule out lung cancer following pneumonia treatment.

4.4 STUDY IV

Background: Overall relative survival in lung cancer has increased in the last few decades. Recent improvements in diagnostic procedures and treatments have possibly affected different subgroups differently.

Study population: Individuals with a first-time diagnosis of adenocarcinoma or squamous cell carcinoma of the lung between 1995 and 2016 as recorded in the NLCR.

Study design: Population-based cohort study.

Setting: Sweden.

Exposure: A diagnosis of lung cancer (ICD-7: 1621, ICD-O-3: C34).

Main outcome: Relative survival at one, two and five years post-diagnosis.

Statistical analyses: Relative survival was estimated for each year between 1995 and 2016. This was performed for the whole cohort, as well as in subgroups defined by sex, histopathology, cancer stage, and smoking status. Survival for the patients with lung cancer was counted from the date of the lung cancer diagnosis until the date of death, emigration or administrative censoring (31 December 2016), whichever occurred first. The expected survival was derived from the general Swedish population using the life table approach.

Results: Among the 36,935 patients diagnosed with adenocarcinoma or squamous cell carcinoma of the lung between 1995 and 2016 in Sweden, one-, two- and five-year age-standardised relative survival increased from 38% (95% CI: 32–43) to 59% (95% CI: 56–61), 21% (95% CI: 17–36) to 37% (95% CI: 35–39), and 14% (95% CI: 10–38) to 24% (95% CI: 21–26) respectively (Figure 6). In general, the estimated relative survival increased most for women (Figure 6), patients with adenocarcinoma (Figure 7), patients with stage III cancer (Figure 8), and never-smokers (Figure 9).

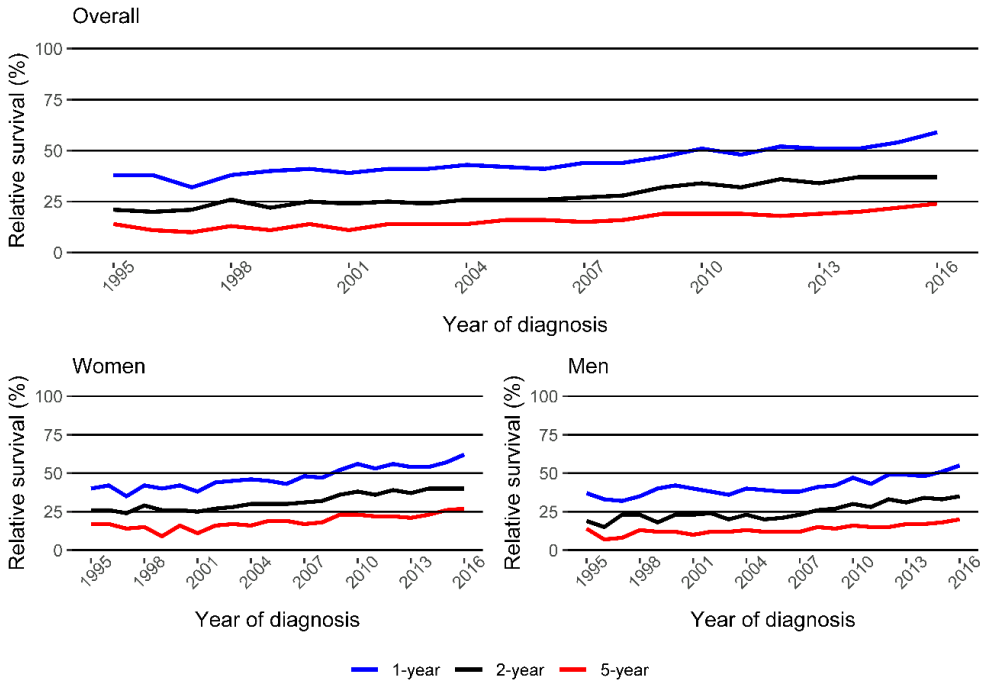

Figure 6. Age-standardised one-, two- and five-year relative survival estimates over calendar years overall and by sex for patients diagnosed with adenocarcinoma or squamous cell carcinoma of the lung, Lung Cancer DataBase Sweden, 1995–2016

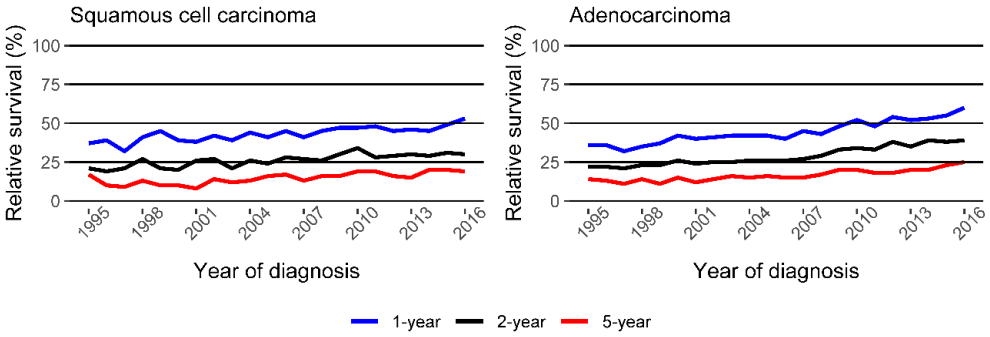

Figure 7. Age-standardised one-, two- and five-year relative survival estimates over calendar years by histopathology for patients diagnosed with lung cancer, Lung Cancer DataBase Sweden, 1995–2016

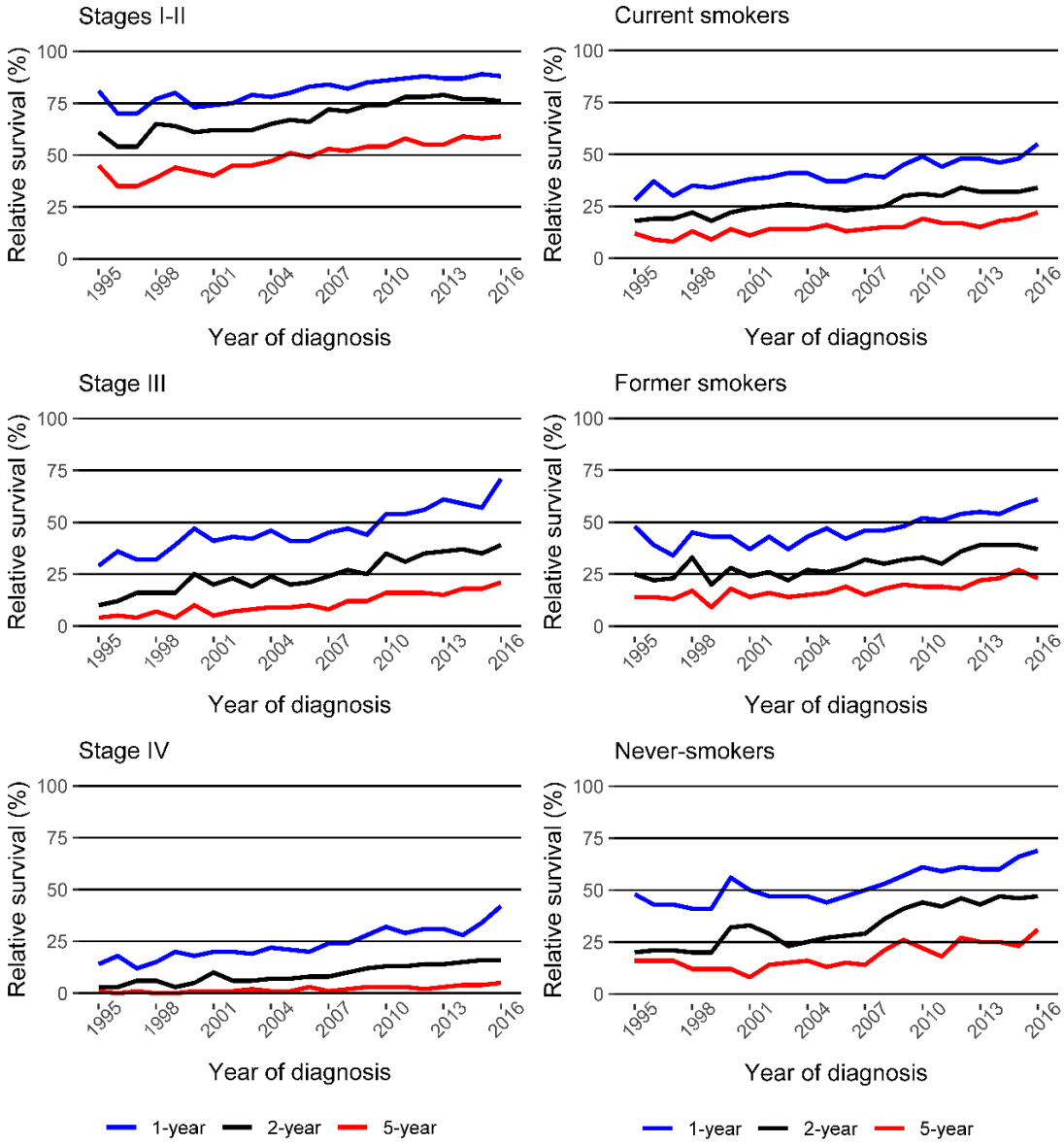

Figure 8. Age-standardised one-, two- and five-year relative survival estimates over calendar years by cancer stage for patients diagnosed with adenocarcinoma or squamous cell carcinoma of the lung, Lung Cancer DataBase Sweden, 1995–2016

Figure 9. Age-standardised one-, two- and five-year relative survival estimates over calendar years by smoking status for patients diagnosed with adenocarcinoma or squamous cell carcinoma of the lung, Lung Cancer DataBase Sweden, 1995–2016

Discussion: The increase in overall five-year relative survival for patients with lung cancer in our study mirrored the trends observed for most countries in the international CONCORD studies for the period 1995–2014 (98, 172).

The more pronounced increase in relative survival that we found for women, patients with adenocarcinoma and never-smokers was not unexpected. In general, patients with targetable mutations are overrepresented among women, patients with adenocarcinoma, and never-smokers (11, 90, 173). Consequently, it should be expected that the introduction of target therapies has been more favourable for these subgroups and has increased the relative survival to a greater extent for them compared to their counterparts.

Confirming recent results from other European countries, we found a more pronounced increase in relative survival for patients with stages I–III compared to stage IV (174-176). The pronounced increase in relative survival for patients with stages I–III may be explained by the fact that the new diagnostic procedures are more frequently used in this group, resulting in more accurate staging and more suitable treatment for these patients (13, 19).

The recent introduction of target therapies and immunotherapies, indicated for patients with late-stage NSCLC, is a probable explanation for some of the estimated increases in relative survival for these patients (19). Even if the indicated increase in five-year relative survival for patients with stage IV cancer was not particularly pronounced in terms of absolute numbers, it must still be considered of special interest. Previously, it was not considered reasonable that improvements for patients with stage IV could affect five-year survival. However, as indicated in our study, this may have started to change, possibly because of the new pharmacological treatments.

Some of the observed increased relative survival for patients with stages III–IV disease in later years may also be attributed to the reclassification of the TNM system in 2010 when patients with pleural effusion were moved from stage IIIB to stage IV (17). This seems to have had a greater impact on relative survival in patients in stage III compared to stage IV.

Conclusion: The results of this study corroborate the previously observed global trend of increased relative survival in patients with lung cancer over time. The relative survival had the most pronounced increase for women, patients with adenocarcinoma, patients with stage III, and never-smokers.

5 METHODOLOGICAL CONSIDERATIONS

5.1 GENERAL CONSIDERATIONS

Selection bias arises when the study population is not representative, in terms of the distribution of exposure status and the outcome, of the underlying source population to which one aims to generalise the results (177). The effect of selection bias on the effect size for the association will differ depending on how the composition of the study population differs from the underlying source population. The use of nationwide population-based registers with high completeness of the entire population reduced the risk of introducing selection bias in the studies included in this thesis.

Misclassification bias arises when the information collected about the study subjects is incorrect (177). This mainly concerns exposure status and the outcome. Misclassification can be differential or non-differential. Non-differential misclassification is a misclassification that is unrelated to other study variables, while for differential misclassification, the degree of misclassification differs according to other variables, for example, exposure status or the outcome. Non-differential misclassification of a binary exposure variable leads to an effect size biased towards the null, namely no association. For a binary outcome variable, non-differential misclassification will bias the effect size towards the null or not introduce any bias. The latter is the case in situations where some individuals with the outcome are incorrectly classified as not having the outcome (outcome sensitivity <100%). Non-differential misclassification of a non-binary variable can bias the effect size towards or away from the null. Differential misclassification can also bias the effect size in either direction.

Due to the high quality of the Swedish population-based registers used in the studies in this thesis, the risk of a substantial impact of misclassification bias was considered to be relatively low. The type of misclassification bias most likely to have been introduced in the studies was non-differential.

Confounding can be defined as a 'confusion of effect' (177), meaning that the effect of the exposure is mixed, or confused, with the effect of another variable, leading to a bias of the estimated effect size. Depending on the association between the confounder and the exposure and the outcome, the effect size can be biased towards or away from the null. A confounder has to be associated with both the exposure and the outcome (Figure 10), cannot be on the causal pathway between the exposure and the outcome (mediator), and cannot be causally influenced by both the exposure and the outcome (collider).

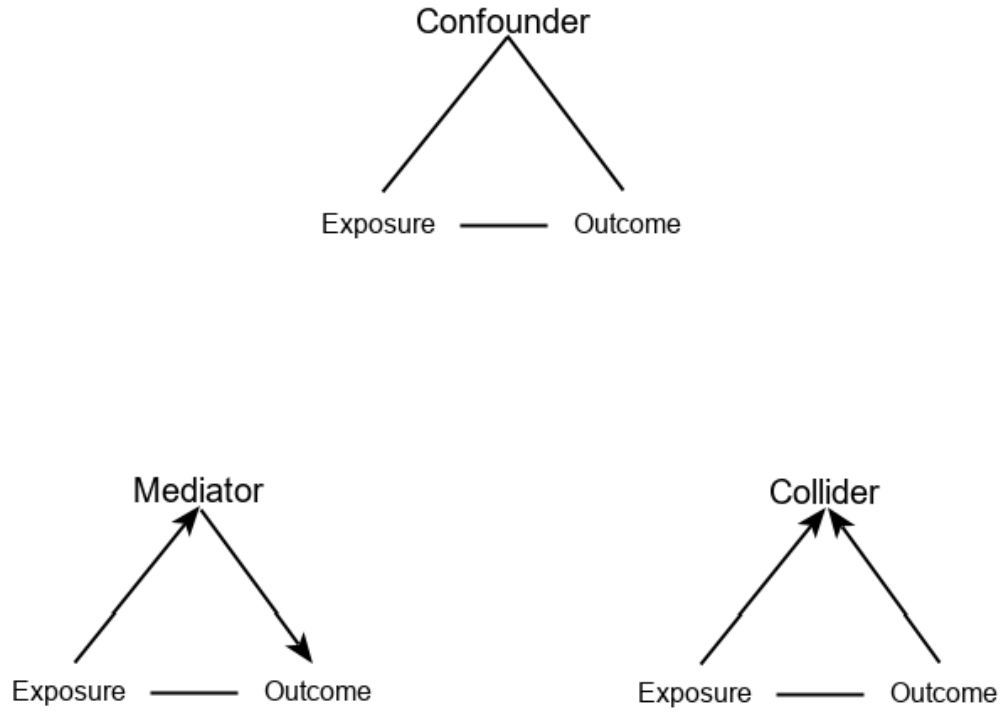

Figure 10. Definition of a confounder

In this thesis, confounding was considered and handled mainly by adjusting for potential confounding factors in the regression models. However, there is always a risk of residual confounding.

5.2 STUDY-SPECIFIC CONSIDERATIONS

5.2.1 Study I

The main limitation in study I was the short follow-up time, as the median follow-up was approximately four years. If there is a true inverse association, one has to consider the time from initiation of the treatment to the effect, mediated by altering the cellular pathways.

Furthermore, we used an intention-to-treat approach, which does not take into account whether exposed individuals stopped taking the medication or whether unexposed individuals started taking the medication after the index date. This would have biased the result towards the null, namely no association. In addition, differences in healthcare-seeking behaviour between unexposed and exposed individuals may have biased the results in study I.

Moreover, the lack of information on potential confounders, for example, smoking status (a diagnosis of chronic obstructive pulmonary disease [COPD] or filled prescription of smoking cessation medication was used as a proxy) and body mass index (BMI) may have resulted in residual confounding.

5.2.2 Study II

Misclassification of exposure status or the outcome are concerns in study II. Possible misclassification was most likely of a non-differential nature, resulting in bias towards the null. However, since the quality and completeness are high in the data sources, a potential bias would only have affected the results in a minor way.

Another limitation is that there was no information in the data source on exposure to passive smoking. The lack of information on passive smoking may have resulted in a biased estimate. Furthermore, we did not have information on changes in smoking status after the diagnosis.

5.2.3 Study III

The lack of information on the indication for the prescribed antibiotics may have introduced misclassification bias of the exposure in study III, namely individuals filled a prescription of an antibiotic that was not prescribed for pneumonia. This was most likely non-differential or more common among individuals free of lung cancer, both of which would have biased the results towards the null. However, since respiratory infections are one of the most common indications for antibiotic use (178), this was likely to have resulted in minor misclassification only. In addition, differences in healthcare-seeking behaviour between the included groups may have biased the results in study III.

Furthermore, the lack of information on potential confounders, for example, smoking status, and exposure to passive smoking and air pollution may have resulted in residual confounding.

5.2.4 Study IV

Assuming a trend of prolonged survival over calendar years, applying the period approach for the most recent years of diagnosis would have underestimated the relative survival for these

years. The assumption of comparability, that the patients would have experienced the same survival as the general population had they been free of the disease of interest, is of the utmost delicacy when applying the relative survival framework. As almost all of the patients with lung cancer are ever-smokers and therefore carry a higher risk of other diseases and consequently a higher risk of all-cause mortality compared to the general population, it has been argued that relative survival should not be used for lung cancer. However, Hinchcliffe *et al* assessed the influence of violating the comparability assumption and concluded that it does not have a concerning impact on the estimated relative survival (179). The prognosis for patients with lung cancer is poor, then the fact that they are also at increased risk of dying from other causes will have little influence on the estimated relative survival.

6 FUTURE PERSPECTIVES

The observed inverse association between use of antimuscarinic medications and lung cancer in study I raises the question of the role of non-cancer therapies in the area of cancer. In times of expensive new cancer therapies, the use of existing non-cancer medications in prevention and treatment of cancer is an interesting and appealing idea. To study and evaluate non-cancer therapies for prevention and treatment of cancer will be an important part of future cancer research.

The differences in lung cancer between patients with and without a history of smoking emphasise the importance of conducting future studies to identify risk factors other than smoking and to reduce the occurrence of these risk factors. The survival difference by smoking history generates questions on underlying reasons that, at least partly, have been addressed in previous studies. Future studies should address how to treat patients in an optimal way based on information on smoking status.

The strong prognostic value of stage in combination with a high proportion of patients with late-stage disease at diagnosis emphasises the importance of methods and strategies for earlier detection and diagnosis. Identifying indicators for early lung cancer that can work as red flags for general practitioners when examining patients should be a focus of future studies. In addition, healthcare planners and providers should consider initiating screening for lung cancer.

As found in study IV, relative survival has increased in recent decades both overall and in subgroups defined by important prognostic factors. However, for some subgroups, the increases were almost negligible. In the future, it will be important to prolong long-term survival for patients with squamous cell carcinoma and stage IV disease as well.

Even if we in the future can make advances in terms of early detection and prolonging survival after diagnosis, the most important action will be to reduce the occurrence of risk factors, especially tobacco smoking. Huge gains will be made in the future if we can prevent the adolescents of today from starting smoking. Therefore, all parts of society have to work together on reducing the prevalence of smoking.

7 POPULÄRVETENSKAPLIG SAMMANFATTNING PÅ SVENSKA

Varje år diagnostiseras ungefär 4000 individer med lungcancer i Sverige och ungefär lika många dör varje år. Lungcancer är den cancerform som i Sverige och globalt orsakar flest dödsfall. Många diagnostiseras i en sen fas av sjukdomen och det är sannolikt starkt bidragande till den höga dödligheten. Att kunna identifiera faktorer som kan minska risken för lungcancer, identifiera tidiga indikatorer på lungcancer, samt att bättre förstå överlevnaden hos patienter med lungcancer och hur överlevnaden har förändrats över tid är av stor vikt för att kunna minska risken för lungcancer men även för att kunna förbättra överlevnaden hos de som får lungcancer.

Resultaten från studie I i avhandlingen indikerar att individer som behandlats med antimuskarina läkemedel mot överaktivblåsa har en lägre sannolikhet att utveckla lungcancer jämfört med individer som inte har erhållit denna form av behandling. Resultaten i studie I genererar tankar och hypoteser om effekter på cancer av läkemedel som inte utvecklats för att behandla cancer.

I studie II såg vi att patienter med lungcancer som aldrig har rökt var äldre, samt att kvinnor, *EGFR* mutation, adenocarcinom och cancer i stadie IV var överrepresenterade hos patienter som aldrig rökt. De som aldrig hade rökt hade också en längre överlevnad efter diagnos jämfört med rökarna. Dessa resultat bidrar till tidigare resultat från andra studier som har sett skillnader mellan rökare och icke-rökare, samt att det understryker vikten av att bättre förstå dessa skillnader för att kunna förebygga lungcancer hos individer som aldrig har rökt och på ett bättre sätt behandla patienter baserat på rökstatus.

På resultaten från studie III ser man att det föreligger ett samband mellan lungcancer och tidigare uthämtning av antibiotika för behandling av lunginfektioner. Styrkan på sambandet ökade med närhet till diagnosen och med antalet uthämtningar under treårsperioden som föregick diagnosen. Dessa resultat antyder att antibiotikaanvändning, speciellt upprepade uttag, kan vara en indikator för tidiga symptom på lungcancer som skulle kunna användas inom primärvården för att kunna upptäcka fler patienter redan i ett tidigt skede.

I studie IV fann vi att den relativa överlevnaden hos patienter med lungcancer har ökat över tid från 1995 till 2016. Den tydligaste ökningen sågs bland patienter med cancer som ännu inte har spritt sig utanför lungorna. Dessa skillnader beror sannolikt på att de förbättringar som gjorts inom diagnostik främst har påverkat patienter i denna grupp, samt att en stor del av dessa patienter även har dragit nytta av nya behandlingar. Kunskapen om hur överlevnaden har förändrats över tid, och skillnaderna i förändring mellan olika grupper, kan bidra till att bättre förstå effekter av genomförda förändringar men även hjälpa till att belysa var framtida insatser bör fokuseras för att framöver kunna förbättra överlevnaden hos flera patienter med lungcancer.

8 ACKNOWLEDGEMENTS

For those of you who went directly to this page, please go back and start on page 1.

Shahram Bahmanyar, *Helle Kieler* and *Gunnar Wagenius*, my supervisors, thank you for taking me on as your doctoral student, for all your support, and for all the scientific input you have given me over the years. I will always remember you.

Thank you to all my co-authors, *Mats Lambe*, *Marie Linder*, *Anders Sundström*, *Fredrik Sandin*, *Annette Karimi* and *Kristina Lamberg*, for all your input on the scientific aspects and the academic writing. You have been an essential part of making this possible.

Thank you to my office mates, *Sarah Burkill* (for all the talks about Brexit and Wildlife Photographer of the Year, for going over the language for most of my work, and for our visits to different pubs around the city, and sorry for eating fruit so loud in the office), *Laura Pazzagli* (my partner in crime when it comes to eating in the office), and *Tatsuya Yagi* and *Aya Nakitanda* (for trying to teach me how to eat with chopsticks – I have not yet mastered that art). You have all made every single day of my time as a doctoral student into an extraordinary first-class journey (by train, not aeroplane, as we still have to think about the climate).

Thank you to all my other colleagues at the Centre for Pharmacoepidemiology, *Lena Brant*, *Philip Brenner*, *Camilla* Byström, C*arolyn Cesta*, *Lotta Ekstrand*, *Karin Gembert*, *Ingvild Odsbu*, *David Hägg*, *Anna Ingemarsdotter*, *Silvia Segovia Chacón*, *Johan Reutfors*, *Zoltan Thinsz*, *Ulrika Undén* and *Caroline Öberg*. Special thanks to *Pär Karlsson* for always taking the time to answer my never-ending questions on statistics, you have definitely been an invaluable source of information.

Thank you to my good friends, *Erik Melander*, *Adam Jarlfors*, *Camilla Johnsson*, *Fredrik Jonsson*, *Ingrid Rask*, *Marcus Forsman*, *Sofia Rhodin*, *Camilo Persson* and *Sebastian Axelsson* for all the good (and not so good) times and for always being there when I need you.

To my father, mother, grandmother, all my siblings (*Rasmus*, *Lotten*, *Jesper*, *Laban* and *Izabella*) and my fiancée, *Elin Gustafsson*, for giving me my moral compass and making me aware of the important things in life.

9 REFERENCES

1. O'Connor CM, Adams JU. *Essentials of Cell Biology*. Cambridge, MA: NPG Education; 2010.

2. Fouad YA, Aanei C. Revisiting the hallmarks of cancer. *Am J Cancer Res* 2017;7(5):1016–1036.

3. Socialstyrelsen – Swedish National Board of Health and Welfare. *Statistics on Cancer Incidence 2018*. Stockholm: Socialstyrelsen – Swedish National Board of Health and Welfare; 2019.

4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68(6):394–424.

5. NORDCAN – Association for the Nordic Cancer Registers. *Cancer Incidence*. https://nordcan.iarc.fr/en (accessed: 2020-08-25).

6. American Cancer Society. *The Cancer Atlas - Lung Cancer 2020*. https://canceratlas.cancer.org/the-burden/lung-cancer/ (accessed: 2020-06-24).

7. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of genetic, clinical and radiologic advances since the 2004 Classification. *J Thorac Oncol* 2015;10(9):1243–1260.

8. Adami H-O, Hunter DJ, Lagiou P, Mucci LA, MacMahon B. *Textbook of Cancer Epidemiology*. New York: Oxford University Press; 2018.

9. Subramanian J, Govindan R. Molecular profile of lung cancer in never smokers. *Eur J Cancer Supplements* 2013;11(2):248–253.

10. Subramanian J, Govindan R. Lung cancer in never smokers: A review. *J Clin Oncol* 2007;25(5):561–570.

11. Donington JS, Colson YL. Sex and gender differences in non-small cell lung cancer. *Semin in Thorac Cardiovasc Surg* 2011;23(2):137–145.

12. American Cancer Society. *What is Lung Cancer?* http://www.cancer.org/content/cancer/en/cancer/lung-cancer/about/what-is.html (accessed: 2020-02-04).

13. Regional Cancer Centre - Uppsala Örebro. *Lungcancer – Nationellkvalitetsrapport för 2018* [*Lung Cancer – National Quality Report for 2018*]. Uppsala: Regional Cancer Centre - Uppsala Örebro; 2019.

14. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 4th ed. Philadelphia: Lippincott; 1992.

15. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 5th ed. Philadelphia: Lippincott-Raven; 1997.

16. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 6th ed. New York: Springer; 2002.

17. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*. 7th ed. New York: Springer; 2010.

18. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American J Clin Oncol* 1982;5(6):649–656.

19. Regionala Cancercentrum i Samverkan. *Nationellt Vårdprogram – Lungcancer* [*National Care Programme – Lung Cancer*]. Uppsala: Regionala Cancercentrum i Samverkan; 2020.

20. Regionala Cancercentrum i Samverkan. *Lungcancer – Standardiserat Vårdförlopp* [*Lung Cancer – Standardised Course of Care*]. Uppsala: Regionala Cancercentrum i Samverkan; 2018.

21. Regionala Cancercentrum i Samverkan. *Väntetid Standardiserat Vårdförlopp - Lungcancer* [*Waiting Time Standardised Course of Care – Lung Cancer*] https://www.cancercentrum.se/samverkan/cancerdiagnoser/lunga-och-lungsack/vardforlopp-lunga/redovisning-vantetid/ (accessed: 2019-11-22).

22. National Health Service. *Lung Cancer – Diagnosis*. https://www.nhs.uk/conditions/lung-cancer/diagnosis/ (accessed: 2019-12-03).

23. Tyczynski JE, Bray F, Parkin DM. Lung cancer in Europe in 2000: Epidemiology, prevention, and early detection. *Lancet Oncol* 2003;4(1):45–55.

24. Parkin DM, Pisani P, Lopez AD, Masuyer E. At least one in seven cases of cancer is caused by smoking. Global estimates for 1985. *Int J Cancer* 1994;59(4):494–504.

25. Cheng T-YD, Cramb SM, Baade PD, Youlden DR, Nwogu C, Reid ME. The international epidemiology of lung cancer: Latest trends, disparities, and tumor characteristics. *J Thorac Oncol* 2016;11(10):1653–1671.

26. Islami F, Torre LA, Jemal A. Global trends of lung cancer mortality and smoking prevalence. *Transl Lung Cancer Res* 2015;4(4):327–338.

27. Smith CJ, Perfetti TA, Rumple MA, Rodgman A, Doolittle DJ. "IARC group 2A Carcinogens" reported in cigarette mainstream smoke. *Food Chem Toxicol* 2000;38(4):371–383.

28. Smith CJ, Perfetti TA, Mullens MA, Rodgman A, Doolittle DJ. "IARC group 2B Carcinogens" reported in cigarette mainstream smoke. *Food Chem Toxicol* 2000;38(9):825–848.

29. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. *BMJ* 1950;2(4682):739–748.

30. Levin ML, Goldstein H, Gerhardt PR. Cancer and tobacco smoking; preliminary report. *JAMA* 1950;143(4):336–338.

31. Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, et al. Tobacco smoking and cancer: A meta-analysis. *Int J Cancer* 2008;122(1):155–164.

32. Hackshaw AK, Law MR, Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. *BMJ* 1997;315(7114):980–988.

33. Canistro D, Vivarelli F, Cirillo S, Babot Marquillas C, Buschini A, Lazzaretti M, et al. E-cigarettes induce toxicological effects that can raise the cancer risk. *Sci Rep* 2017;7(1):2028.

34. Shields PG, Berman M, Brasky TM, Freudenheim JL, Mathe E, McElroy JP, et al. A review of pulmonary toxicity of electronic cigarettes in the context of smoking: A focus on inflammation. *Cancer Epidemiol Biomarkers Prev* 2017;26(8):1175–1191.

35. Boffetta P, Hecht S, Gray N, Gupta P, Straif K. Smokeless tobacco and cancer. *Lancet Oncol*. 2008;9(7):667–675.

36. Sethi TK, El-Ghamry MN, Kloecker GH. Radon and lung cancer. *Clin Adv Hematol Oncol* 2012;10(3):157–164.

37. Samet JM. Radiation and cancer risk: A continuing challenge for epidemiologists. *Environmental Health* 2011;10(Suppl 1):S4.

38. Pershagen G, Akerblom G, Axelson O, Clavensjo B, Damber L, Desai G, et al. Residential radon exposure and lung cancer in Sweden. *New Eng J Med* 1994;330(3):159–164.

39. Lagarde F, Axelsson G, Damber L, Mellander H, Nyberg F, Pershagen G. Residential radon and lung cancer among never-smokers in Sweden. *Epidemiol* 2001;12(4):396–404.

40. Prazakova S, Thomas PS, Sandrini A, Yates DH. Asbestos and the lung in the 21st century: An update. *Clin Resp J* 2014;8(1):1–10.

41. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: Epidemiology, etiology, and prevention. *Clin Chest Med* 2011;32(4):605–644.

42. World Health Organization. *Asbestos: Elimination of Asbestos-related Diseases* http://www.who.int/news-room/fact-sheets/detail/asbestos-elimination-of-asbestos-related-diseases (accessed: 2018-06-12).

43. Lynch KM, Smith WA. Pulmonary asbestosis: V. A report of bronchial carcinoma and epithelial metaplasia. *Am J Can*, 1939;36(4):567–573.

44. Loomis D, Huang W, Chen G. The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: Focus on China. *Chin J Can* 2014;33(4):189–196.

45. Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol* 2013;14(9):813–22.

46. Liang HY, Li XL, Yu XS, Guan P, Yin ZH, He QC, et al. Facts and fiction of the relationship between preexisting tuberculosis and lung cancer risk: A systematic review. *Int J Cancer* 2009;125(12):2936–2944.

47. Littman AJ, Jackson LA, Vaughan TL. Chlamydia pneumoniae and lung cancer: Epidemiologic evidence. *Cancer Epidemiol Biomarkers Prev* 2005;14(4):773–778.

48. Dalton-Griffin L, Kellam P. Infectious causes of cancer and their detection. *J Biol.* 2009;8(7):67.

49. Hicks BM, Filion KB, Yin H, Sakr L, Udell JA, Azoulay L. Angiotensin converting enzyme inhibitors and risk of lung cancer: Population based cohort study. *BMJ* 2018;363:k4209.

50. Zhang H, Garcia Rodriguez LA, Hernandez-Diaz S. Antibiotic use and the risk of lung cancer. *Cancer Epidemiol Biomarkers Prev* 2008;17(6):1308–1315.

51. Boursi B, Mamtani R, Haynes K, Yang YX. Recurrent antibiotic exposure may promote cancer formation--Another step in understanding the role of the human microbiota? *Eur J Cancer* 2015;51(17):2655–2664.

52. Maddi A, Sabharwal A, Violante T, Manuballa S, Genco R, Patnaik S, et al. The microbiome and lung cancer. *J Thorac Dis* 2019;11(1):280–291.

53. Oh SW, Myung SK, Park JY, Lee CM, Kwon HT. Aspirin use and risk for lung cancer: A meta-analysis. *Ann Oncol* 2011;22(11):2456–2465.

54. World Cancer Research Fund/American Institute for Cancer Research. *Diet, Nutrition, Physical Activity and Lung Cancer*. https://www.wcrf.org/sites/default/files/Lung-cancer-report.pdf (accessed: 2020-08-31).

55. Cote ML, Liu M, Bonassi S, Neri M, Schwartz AG, Christiani DC, et al. Increased risk of lung cancer in individuals with a family history of the disease: A pooled analysis from the International Lung Cancer Consortium. *Eur J Cancer* 2012;48(13):1957–1968.

56. Cook MB, Dawsey SM, Freedman ND, Inskip PD, Wichner SM, Quraishi SM, et al. Sex disparities in cancer incidence by period and age. *Cancer Epidemiol Biomarkers Prev* 2009;18(4):1174–1182.

57. Edgren G, Liang L, Adami HO, Chang ET. Enigmatic sex disparities in cancer incidence. *Eur J Epidemiol* 2012;27(3):187–196.

58. Radkiewicz C, Johansson ALV, Dickman PW, Lambe M, Edgren G. Sex differences in cancer risk and survival: A Swedish cohort study. *Eur J Cancer* 2017;84:130–140.

59. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol* 2016;16(10):626–638.

60. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC). ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017;28(suppl_4):iv1–iv21.

61. Graham EA, Singer JJ. Successful removal of an entire lung for carcinoma of the bronchus. *JAMA* 1933;101(18):1371–1374.

62. Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung adjuvant cisplatin evaluation: A pooled analysis by the LACE Collaborative Group. *J Clin Oncol* 2008;26(21):3552–3559.

63. Arriagada R, Auperin A, Burdett S, Higgins JP, Johnson DH, Le Chevalier T, et al. Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small-cell lung cancer: Two meta-analyses of individual patient data. *Lancet* 2010;375(9722):1267–1277.

64. Scagliotti GV, Pastorino U, Vansteenkiste JF, Spaggiari L, Facciolo F, Orlowski TM, et al. Randomized phase III study of surgery alone or surgery plus preoperative cisplatin and gemcitabine in stages IB to IIIA non-small-cell lung cancer. *J Clin Oncol* 2011;30(2):172–178.

65. MacLean M, Luo X, Wang S, Kernstine K, Gerber DE, Xie Y. Outcomes of neoadjuvant and adjuvant chemotherapy in stage 2 and 3 non-small cell lung cancer: An analysis of the National Cancer Database. *Oncotarget* 2018;9(36):24470–24479.

66. Nyman J, Hallqvist A, Lund JA, Brustugun OT, Bergman B, Bergstrom P, et al. SPACE – A randomized study of SBRT vs conventional fractionated radiotherapy in medically inoperable stage I NSCLC. *Radiother Oncol* 2016;121(1):1–8.

67. Haque W, Verma V, Polamraju P, Farach A, Butler EB, Teh BS. Stereotactic body radiation therapy versus conventionally fractionated radiation therapy for early stage non-small cell lung cancer. *Radiother Oncol* 2018;129(2):264–269.

68. Jassem J. Combined chemotherapy and radiation in locally advanced non-small-cell lung cancer. *Lancet Oncol* 2001;2(6):335–342.

69. Auperin A, Le Pechoux C, Rolland E, Curran WJ, Furuse K, Fournel P, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *J Clin Oncol* 2010;28(13):2181–2190.

70. O'Rourke N, Roque IFM, Farre Bernado N, Macbeth F. Concurrent chemoradiotherapy in non-small cell lung cancer. *The Cochrane Database of Systematic Reviews* 2010(6):Cd002140.

71. Sgambato A, Casaluce F, Sacco PC, Palazzolo G, Maione P, Rossi A, et al. Anti PD-1 and PDL-1 immunotherapy in the treatment of advanced non-small cell lung cancer (NSCLC): A review on toxicity profile and its management. *Curr Drug Saf* 2016;11(1):62–68.

72. NSCLC Meta-Analyses Collaborative Group. Chemotherapy in addition to supportive care improves survival in advanced non-small-cell lung cancer: A systematic review and meta-analysis of individual patient data from 16 randomized controlled trials. *J Clin Oncol* 2008;26(28):4617–4625.

73. Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: A meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ* 1995;311(7010):899–909.

74. Ardizzoni A, Boni L, Tiseo M, Fossella FV, Schiller JH, Paesmans M, et al. Cisplatin-versus carboplatin-based chemotherapy in first-line treatment of advanced non-small-cell lung cancer: An individual patient data meta-analysis. *J Nat Cancer Inst* 2007;99(11):847–857.

75. Soon YY, Stockler MR, Askie LM, Boyer MJ. Duration of chemotherapy for advanced non-small-cell lung cancer: A systematic review and meta-analysis of randomized trials. *J Clin Oncol* 2009;27(20):3277–3283.

76. Lima JP, dos Santos LV, Sasse EC, Sasse AD. Optimal duration of first-line chemotherapy for advanced non-small cell lung cancer: A systematic review with meta-analysis. *Eur J Cancer* 2009;45(4):601–607.

77. Rossi A, Chiodini P, Sun JM, O'Brien ME, von Plessen C, Barata F, et al. Six versus fewer planned cycles of first-line platinum-based chemotherapy for non-small-cell lung cancer: A systematic review and meta-analysis of individual patient data. *Lancet Oncol* 2014;15(11):1254–1262.

78. Chen DS, Irving BA, Hodi FS. Molecular pathways: Next-generation immunotherapy--inhibiting programmed death-ligand 1 and programmed death-1. *Clin Cancer Res* 2012;18(24):6580–6587.

79. Reck M, Rodriguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fulop A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *New Eng J Med* 2016;375(19):1823–1833.

80. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *New Eng J Med* 2018;378(22):2078–2092.

81. West H, McCleod M, Hussein M, Morabito A, Rittmeyer A, Conter HJ, et al. Atezolizumab in combination with carboplatin plus nab-paclitaxel chemotherapy compared with chemotherapy alone as first-line treatment for metastatic non-squamous non-small-cell lung cancer (IMpower130): A multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2019;20(7):924-937.

82. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 2007;7(3):169–181.

83. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *New Eng J Med* 2010;362(25):2380–2388.

84. Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): A multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;12(8):735–742.

85. Gao G, Ren S, Li A, Xu J, Xu Q, Su C, et al. Epidermal growth factor receptor-tyrosine kinase inhibitor therapy is effective as first-line treatment of advanced non-small-cell lung cancer with mutated EGFR: A meta-analysis from six phase III randomized controlled trials. *Int J Cancer* 2012;131(5):822–829.

86. Yang JC, Wu YL, Schuler M, Sebastian M, Popat S, Yamamoto N, et al. Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): Analysis of overall survival data from two randomised, phase 3 trials. *Lancet Oncol* 2015;16(2):141–151.

87. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *New Eng J Med* 2009;361(10):947–957.

88. Hosomi Y, Morita S, Sugawara S, Kato T, Fukuhara T, Gemma A, et al. Gefitinib alone versus gefitinib plus chemotherapy for non-small-cell lung cancer with mutated epidermal growth factor receptor: NEJ009 study. *J Clin Oncol* 2019;38(2):115–123.

89. Huang Q, Li J, Sun Y, Wang R, Cheng X, Chen H. Efficacy of EGFR tyrosine kinase inhibitors in the adjuvant treatment for operable non-small cell lung cancer by a meta-analysis. *Chest* 2016;149(6):1384–1392.

90. Du X, Shao Y, Qin H-F, Tai Y-H, Gao H-J. ALK-rearrangement in non-small-cell lung cancer (NSCLC). *Thorac Cancer* 2018;9(4):423–430.

91. Solomon BJ, Mok T, Kim DW, Wu YL, Nakagawa K, Mekhail T, et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *New Eng J Med* 2014;371(23):2167–2177.

92. Soria JC, Tan DSW, Chiari R, Wu YL, Paz-Ares L, Wolf J, et al. First-line ceritinib versus platinum-based chemotherapy in advanced ALK-rearranged non-small-cell lung cancer (ASCEND-4): A randomised, open-label, phase 3 study. *Lancet* 2017;389(10072):917–929.

93. European Medicines Agency. *Avastin: EPAR - Product Information*. London: European Medicines Agency; 2019.

94. Johnson DH, Fehrenbacher L, Novotny WF, Herbst RS, Nemunaitis JJ, Jablons DM, et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. *J Clin Oncol* 2004;22(11):2184–2191.

95. Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *New Eng J Med* 2006;355(24):2542–2550.

96. Niho S, Kunitoh H, Nokihara H, Horai T, Ichinose Y, Hida T, et al. Randomized phase II study of first-line carboplatin-paclitaxel with or without bevacizumab in Japanese patients with advanced non-squamous non-small-cell lung cancer. *Lung Cancer* 2012;76(3):362–367.

97. Reck M, von Pawel J, Zatloukal P, Ramlau R, Gorbounova V, Hirsh V, et al. Phase III trial of cisplatin plus gemcitabine with either placebo or bevacizumab as first-line therapy for nonsquamous non-small-cell lung cancer: AVAil. *J Clin Oncol* 2009;27(8):1227–1234.

98. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): Analysis of individual

records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;391(10125):1023–1075.

99. Clark GM, Zborowski DM, Culbertson JL, Whitehead M, Savoie M, Seymour L, et al. Clinical utility of epidermal growth factor receptor expression for selecting patients with advanced non-small cell lung cancer for treatment with erlotinib. *J Thorac Oncol* 2006;1(8):837–846.

100. Paesmans M. Prognostic and predictive factors for lung cancer. *Breathe* 2012;9(2):112–121.

101. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WEE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 2016;11(1):39–51.

102. Cuyun Carter G, Barrett AM, Kaye JA, Liepa AM, Winfree KB, John WJ. A comprehensive review of nongenetic prognostic and predictive factors influencing the heterogeneity of outcomes in advanced non-small-cell lung cancer. *Cancer Manag Res* 2014;6:437–449.

103. Steels E, Paesmans M, Berghmans T, Branle F, Lemaitre F, Mascaux C, et al. Role of p53 as a prognostic factor for survival in lung cancer: A systematic review of the literature with a meta-analysis. *Eur Respir J* 2001;18(4):705–719.

104. Meert AP, Martin B, Delmotte P, Berghmans T, Lafitte JJ, Mascaux C, et al. The role of EGF-R expression on patient survival in lung cancer: A systematic review with meta-analysis. *Eur Respir J* 2002;20(4):975–981.

105. Meng D, Yuan M, Li X, Chen L, Yang J, Zhao X, et al. Prognostic value of K-RAS mutations in patients with non-small cell lung cancer: A systematic review with meta-analysis. *Lung Cancer* 2013;81(1):1–10.

106. Kawaguchi T, Takada M, Kubo A, Matsumura A, Fukai S, Tamura A, et al. Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: A comprehensive analysis of 26,957 patients with NSCLC. *J Thorac Oncol* 2010;5(5):620–630.

107. Nakamura H, Ando K, Shinmyo T, Morita K, Mochizuki A, Kurimoto N, et al. Female gender is an independent prognostic factor in non-small-cell lung cancer: A meta-analysis. *Ann Thorac Cardiovasc Surg* 2011;17(5):469–480.

108. Wisnivesky JP, Halm EA. Sex differences in lung cancer survival: Do tumors behave differently in elderly women? *J Clin Oncol* 2007;25(13):1705–1712.

109. Van der Heyden JHA, Schaap MM, Kunst AE, Esnaola S, Borrell C, Cox B, et al. Socioeconomic inequalities in lung cancer mortality in 16 European populations. *Lung Cancer* 2009;63(3):322–330.

110. Clement-Duchene C, Stock S, Xu X, Chang ET, Gomez SL, West DW, et al. Survival among never-smokers with lung cancer in the Cancer Care Outcomes Research and Surveillance Study. *Ann Am Thorac Soc* 2016;13(1):58–66.

111. Kawaguchi T, Matsumura A, Fukai S, Tamura A, Saito R, Zell JA, et al. Japanese ethnicity compared with Caucasian ethnicity and never-smoking status are independent favorable prognostic factors for overall survival in non-small cell lung cancer: A collaborative epidemiologic study of the National Hospital Organization Study Group for Lung Cancer (NHSGLC) in Japan and a Southern California Regional Cancer Registry databases. *J Thorac Oncol* 2010;5(7):1001–1010.

112. Yano T, Miura N, Takenaka T, Haro A, Okazaki H, Ohba T, et al. Never-smoking nonsmall cell lung cancer as a separate entity: Clinicopathologic features and survival. *Cancer* 2008;113(5):1012–1018.

113. Willén L, Berglund A, Bergström S, Bergqvist M, Öjdahl-Bodén A, Wagenius G, et al. Educational level and management and outcomes in non-small cell lung cancer. A nationwide population-based study. *Lung Cancer* 2019;131:40–46.

114. Berglund A, Lambe M, Luchtenborg M, Linklater K, Peake MD, Holmberg L, et al. Social differences in lung cancer management and survival in South East England: A cohort study. *BMJ Open* 2012;2(3).

115. Parsons A, Daley A, Begh R, Aveyard P. Influence of smoking cessation after diagnosis of early stage lung cancer on prognosis: Systematic review of observational studies with meta-analysis. *BMJ* 2010;340:b5569.

116. Dresler CM. Is it more important to quit smoking than which chemotherapy is used? *Lung Cancer* 2003;39(2):119–124.

117. O'Malley M, King AN, Conte M, Ellingrod VL, Ramnath N. Effects of cigarette smoking on metabolism and effectiveness of systemic therapy for lung cancer. *J Thorac Oncol* 2014;9(7):917–926.

118. Tammemagi CM, Neslund-Dudas C, Simoff M, Kvale P. Smoking and lung cancer survival: The role of comorbidity and treatment. *Chest* 2004;125(1):27–37.

119. Sorensen LT, Horby J, Friis E, Pilsgaard B, Jorgensen T. Smoking as a risk factor for wound healing and infection in breast cancer surgery. *Eur J Surg Oncol* 2002;28(8):815–820.

120. Ludvigsson JF, Almqvist C, Bonamy AE, Ljung R, Michaelsson K, Neovius M, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016;31(2):125–136.

121. Regional Cancer Centre – Uppsala Örebro. *Swedish National Lung Cancer Register 2017* http://www.cancercentrum.se/vast/cancerdiagnoser/lunga-och-lungsack/kvalitetsregister/ (accessed: 2017-09-13).

122. Socialstyrelsen – Swedish National Board of Health and Welfare. *The Swedish Cancer Register 2019* https://www.socialstyrelsen.se/en/statistics-and-data/registers/register-information/swedish-cancer-register/ (accessed: 2019-11-22).

123. Socialstyrelsen – Swedish National Board of Health and Welfare. *The Swedish Cause of Death Register* https://www.socialstyrelsen.se/statistik-och-data/register/alla-register/dodsorsaksregistret/ (accessed: 2019-11-22).

124. Socialstyrelsen – Swedish National Board of Health and Welfare. *The Swedish National Patient Register* https://www.socialstyrelsen.se/en/statistics-and-data/registers/register-information/the-national-patient-register/ (accessed: 2019-11-22).

125. Socialstyrelsen – Swedish National Board of Health and Welfare. *The Swedish Prescribed Drug Register* https://www.socialstyrelsen.se/statistik-och-data/register/alla-register/lakemedelsregistret/ (accessed: 2019-11-22).

126. Statistiska Centralbyrån (SCB) – Statistics Sweden. Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA) https://www.scb.se/contentassets/f0bc88c852364b6ea5c1654a0cc90234/dokumentation-av-lisa.pdf (accessed: 2017-10-26).

127. Skatteverket – The Swedish Tax Agency. *The Swedish Population Register* https://www.skatteverket.se/privat/folkbokforing/attvarafolkbokford/folkbokforingsdatabasen.4.3810a01c150939e893f16fe2.html (accessed: 2020-06-04).

128. Regional Cancer Centre – Uppsala Örebro. *Swedish National Lung Cancer Register - Inclusion Criteria and Diagnosis Selection*. Uppsala: Regional Cancer Centre – Uppsala Örebro; 2017.

129. Regional Cancer Centre – Uppsala Örebro. *Styrdokument Nationella Lungcancerregistret och Mesoteliomregistret 2018*. Uppsala: Regional Cancer Centre – Uppsala Örebro; 2018

130. Socialstyrelsen – Swedish National Board of Health and Welfare. *Kodning i Cancerregistret – Handledning 2015*. Stockholm: Socialstyrelsen – Swedish National Board of Health and Welfare; 2015.

131. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: A sample survey for year 1998. *Acta Oncol* 2009;48(1):27–33.

132. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry--history, content, quality and use. *Danish Medical Bulletin* 1997;44(5):535–539.

133. Brewster DH, Crichton J, Harvey JC, Dawson G. Completeness of case ascertainment in a Scottish regional cancer registry for the year 1992. *Public Health* 1997;111(5):339–343.

134. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncol* 1994;33(4):365–369.

135. Brooke HL, Talback M, Hornblad J, Johansson LA, Ludvigsson JF, Druid H, et al. The Swedish Cause of Death Register. *Eur J Epidemiol* 2017;32(9):765–773.

136. Johansson LA, Bjorkenstam C, Westerling R. Unexplained differences between hospital and mortality data indicated mistakes in death certification: An investigation of 1,094 deaths in Sweden during 1995. *J Clin Epidemiol* 2009;62(11):1202–1209.

137. World Health Organization. *International statistical classification of diseases and related health problems – 10th revision*. Geneva: World Health Organization; 2010.

138. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.

139. Wettermark B, Hammar N, Fored CM, Leimanis A, Otterblad Olausson P, Bergman U, et al. The new Swedish Prescribed Drug Register--opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007;16(7):726–735.

140. Kartsonaki C. Survival analysis. *Diagn Histopathol* 2016;22(7):263–270.

141. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53(282):457–481.

142. Friedman LM, Furberg C, DeMets D, Reboussin, D, Granger C. *Fundamentals of clinical trials.*New York: Springer; 2015.

143. Cox DR. Regression models and life-tables. *J Royal Stat Soc* 1972;34(2):187–220.

144. Schaffar R, Rachet B, Belot A, Woods LM. Estimation of net survival for cancer patients: Relative survival setting more robust to some assumption violations than cause-specific setting, a sensitivity analysis on empirical data. *Eur J Cancer* 2017;72:78–83.

145. Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: Elements for further discussion. *Stat Med* 1990;9(5):529–538.

146. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: A statistical methodology. *National Cancer Institute Monograph* 1961;6:101–121.

147. Pohar Perme M, Esteve J, Rachet B. Analysing population-based cancer survival – settling the controversies. *BMC Cancer* 2016;16(1):933.

148. Hallas J, Margulis AV, Pottegard A, Kristiansen NS, Atsma WJ, Appenteng K, et al. Incidence of common cancers in users of antimuscarinic medications for overactive bladder: A Danish nationwide cohort study. *Basic Clin Pharmacol Toxicol* 2018;122(6):612-619.

149. Radzikowska E, Glaz P, Roszkowski K. Lung cancer in women: Age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases. *Ann Oncol* 2002;13(7):1087–1093.

150. Pelosof L, Ahn C, Gao A, Horn L, Madrigales A, Cox J, et al. Proportion of never-smoker non-small cell lung cancer patients at three diverse institutions. *J Nat Cancer Inst* 2017;109(7):djw295.

151. Huang Y, Wang R, Pan Y, Zhang Y, Li H, Cheng C, et al. Clinical and genetic features of lung squamous cell cancer in never-smokers. *Oncotarget* 2016;7(24):35979–35988.

152. Dias M, Linhas R, Campainha S, Conde S, Barroso A. Lung cancer in never-smokers – what are the differences? *Acta Oncol* 2017;56(7):931–935.

153. Bryant A, Cerfolio RJ. Differences in epidemiology, histology, and survival between cigarette smokers and never-smokers who develop non-small cell lung cancer. *Chest* 2007;132(1):185–192.

154. Nordquist LT, Simon GR, Cantor A, Alberts WM, Bepler G. Improved survival in never-smokers vs current smokers with primary adenocarcinoma of the lung. *Chest* 2004;126(2):347–351.

155. Pallis AG, Syrigos KN. Lung cancer in never smokers: Disease characteristics and risk factors. *Crit Rev Oncol Hematol* 2013;88(3):494–503.

156. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. *Nat Rev Cancer* 2007;7(10):778–790.

157. Toh CK, Lim WT. Lung cancer in never-smokers. *J Clin Pathol* 2007;60(4):337–340.

158. Toh CK, Gao F, Lim WT, Leong SS, Fong KW, Yap SP, et al. Never-smokers with lung cancer: Epidemiologic evidence of a distinct disease entity. *J Clin Oncol* 2006;24(15):2245–2251.

159. Toh CK, Wong EH, Lim WT, Leong SS, Fong KW, Wee J, et al. The impact of smoking status on the behavior and survival outcome of patients with advanced non-small cell lung cancer: A retrospective analysis. *Chest* 2004;126(6):1750–1756.

160. Clement-Duchene C, Vignaud JM, Stoufflet A, Bertrand O, Gislard A, Thiberville L, et al. Characteristics of never smoker lung cancer including environmental and occupational risk factors. *Lung Cancer* 2010;67(2):144–150.

161. Zheng S, Wang R, Zhang Y, Pan Y, Cheng C, Zheng D, et al. Former smokers with non-small-cell lung cancers: A comprehensive investigation of clinicopathologic characteristics, oncogenic drivers, and prognosis. *Cancer Med* 2016;5(8):2117–2125.

162. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. Mutations of the strong epidermal growth factor receptor gene in lung cancer. *Biol Clin Imp* 2004;64(24):8919–8923.

163. Motono N, Funasaki A, Sekimura A, Usuda K, Uramoto H. Prognostic value of epidermal growth factor receptor mutations and histologic subtypes with lung adenocarcinoma. *Med Oncol* 2018;35(3):22.

164. Rivera GA, Wakelee H. Lung cancer in never smokers. *Adv Exp Med Biol* 2016;893:43–57.

165. Wakelee HA, Chang ET, Gomez SL, Keegan TH, Feskanich D, Clarke CA, et al. Lung cancer incidence in never smokers. *J Clin Oncol* 2007;25(5):472–478.

166. Brandt A, Bermejo JL, Sundquist J, Hemminki K. Age of onset in familial cancer. *Ann Oncol* 2008;19(12):2084–2088.

167. Cassim S, Chepulis L, Keenan R, Kidd J, Firth M, Lawrenson R. Patient and carer perceived barriers to early presentation and diagnosis of lung cancer: A systematic review. *BMC Cancer* 2019;19(1):25.

168. Kölbeck K-G, Almer H. *Regionalt Vårdprogram – Lungcancer* [Regional Care Programme – Lung Cancer] http://www.viss.nu/Handlaggning/Vardprogram/Andning/Lungcancer/ (accessed: 2019-01-14).

169. National Health Service. *Lung Cancer Symptoms*. https://www.nhs.uk/conditions/lung-cancer/symptoms/# (accessed: 2019-08-27).

170. Bjerager M, Palshof T, Dahl R, Vedsted P, Olesen F. Delay in diagnosis of lung cancer in general practice. *Br J Gen Pract* 2006;56(532):863–868.

171. Stralin K, Goscinski G, Hedlund J, Lidman C, Spindler C, Ortqvist A, et al. Management of adult patients with community-acquired pneumonia. Evidence-based guidelines from the Swedish Infectious Diseases Association. *Lakartidningen* 2008;105(38):2582–2587.

172. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang XS, et al. Global surveillance of cancer survival 1995–2009: Analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 2015;385(9972):977–1010.

173. Isaksson S. *Blood- and tumour-based analyses for improved prognostics in lung cancer* [dissertation]. Lund: Lund University; 2019.

174. Driessen EJ, Aarts MJ, Bootsma GP, van Loon JG, Janssen-Heijnen ML. Trends in treatment and relative survival among non-small cell lung cancer patients in the Netherlands (1990–2014): Disparities between younger and older patients. *Lung Cancer* 2017;108:198–204.

175. Innos K, Oselin K, Laisaar T, Aareleid T. Patterns of survival and surgical treatment in lung cancer patients in Estonia by histologic type and stage, 1996–2016. *Acta Oncol* 2019;58(11):1549-1556.

176. Nilssen Y, Strand TE, Fjellbirkeland L, Bartnes K, Møller B. Lung cancer survival in Norway, 1997–2011: From nihilism to optimism. *Eur Respir J* 2016;47(1):275–287.

177. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia: Wolters Kluwer; 2008.

178. Aabenhus R, Hansen MP, Siersma V, Bjerrum L. Clinical indications for antibiotic use in Danish general practice: Results from a nationwide electronic prescription database. S*cand J Prim Health Care* 2017;35(2):162–169.

179. Hinchliffe SR, Rutherford MJ, Crowther MJ, Nelson CP, Lambert PC. Should relative survival be used with lung cancer data? *Br J Cancer* 2012;106:1854–1859.